INTERNET PROTOCOL (IP) NETWORK MEASUREMENT, CHARACTERIZATION, MODELING, AND CONTROL FOR SELF-MANAGED NETWORKS

Purdue University

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2005-341 has been reviewed and is approved for publication

APPROVED: /s/

GREGORY HADYNSKI
Project Engineer

FOR THE DIRECTOR: /s/

WARREN H. DEBANY, JR., Technical Advisor
Information Grid Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE SEPTEMBER 2005	3. REPORT TYPE AND DATES COVERED Final Apr 04 – Apr 05

4. TITLE AND SUBTITLE
INTERNET PROTOCOL (IP) NETWORK MEASUREMENT, CHARACTERIZATION, MODELING, AND CONTROL FOR SELF-MANAGED NETWORKS

6. AUTHOR(S)
William S. Cleveland, Hui Chen,
Bowei Xi and Jin Cao

5. FUNDING NUMBERS
C   - FA8750-04-1-0150
PE   - 62301E
PR   - N162
TA   - PU
WU   - RD

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Purdue University
Statistics Department
150 North University Street
West Lafayette Indiana 47907-2067

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency   AFRL/IFGC
3701 North Fairfax Drive   525 Brooks Road
Arlington Virginia 22203-1714   Rome New York 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2005-341

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Gregory Hadynski/IFGC/(315) 330-4094/ Gregory.Hadynski@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*
IP network technology cannot continue on an ever-increasing course of technological complexity and yet require the kind of human intervention that is necessary today for network management. Networks must be self-managing. This can only be done by a system of measurement that copes with the dynamics of packet movement. This system must process the packet-level measurements into variables that characterize network behaviors, which then form the basis for control algorithms that react to the variables. Such packet level measurements can lead to characterization and control at the application layer, at the transport layer, at the network layer (including overlay networking), and in some cases at the link layer. The research documented in this report used tools of statistics, data mining and machine learning to (1) determine network variables derivable from the measurements that characterize network behavior; (2) develop models of the critical network variables that characterize performance, usage, security, and early onset of problems; and (3) develop automated control methods based on the variables.

14. SUBJECT TERMS
Network Measurement, Network Characterization, Network Modeling, Network Control

15. NUMBER OF PAGES
54

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Contents

List of Figures

1   Overview: Introduction

IP network technology cannot continue on an ever-increasing course o technological complexity and yet require the kind of human intervention that is necessary today for network management. Networks must be self-managing. This can only be done by a system of measurement that copes with the dynamics of packet movement. This system must process the packet-level measurements into variables that characterize network behaviors, which then form the basis for control algorithms that react to the variables. Such packet-level measurements can lead to characterization and control at the application layer, at the transport layer, at the network layer (including overlay networking), and in some cases at the link layer. This research will use tools of statistics, data mining and machine learning to explore packet-level measurement for self-management.

The current dominant measurement framework on IP networks, SNMP, is not adequate to the task because it does not allow the determination of network time dynamics. We studied data that allows the determination of time dynamics—time-stamping the arrivals of packets at network nodes and recording header contents. The research used methods of statistics, data mining, and machine learning to:

(1) determine network variables derivable from the measurements that characterize network behavior;

(2) develop models of the critical network variables that characterize performance, usage, security, and early onset of problems; and

(3) develop automated control methods based on the variables. In addition, certain models that we developed can provide traffic generation in network simulation systems

that can be used for further study of control.

Our project involved the study of previously collected packet header data on IP network links, as well as providing for collection of new data.

The statistics, data mining, and machine learning was carried out using the S-Net software system for IP network packet-level data. S-Net, the most comprehensive environment for packet header study available today, was developed with support, in part, by DARPA under federal contract No. F30602-00-C-0034.

This report describes the work done under the above federal contract as well as previous work done under federal contract No. F30602-02-2-C0093 awarded to Bell Labs because the work under the new contract was a continuation of the work under the previous contract. In a later section of this final report the part of the work done under the new contract is described.

2   Overview: Main Elements of the Approach

Assembled a database of packet header data. Partnered with network operators who were a source not only of data, but provided knowledge of the critical network operations issues. Partnered with network equipment manufacturers to determine feasiblity of different packet-level measurement frameworks. Statistical analysis, data mining, and machine learning used to study network behavior using the S-Net system for packet-header analysis and visualization. Developed models for the measurements. Developed algorithms for self-management based on the measurement scenario, models, and the learned network behavior.

3   Overview: Deliverables

1. *Management and administration:* As part of the conventional DARPA monthly reporting process, project management deliverables included the status of the effort, progress toward accomplishment of contract requirements, and monitoring funding use rates (spreadsheet format for financial reports with actual monthly and cumulative expenditures compared with the programmed budget). We also made oral presentations at two DARPA PI meetings.

2. *Research:* Results on measurement frameworks, network characterization, models, variables that characterize network behavior, and self-managing control methods based on the variables, conveyed through viewgraphs and written documents.

4   Overview: Statement of Work

We collaborated and cooperated with the other program contractors on the planning, design, implementation and demonstration of new technologies described within this proposal.

Over the course of execution of the contract resulting from this proposal, we maintained effective communications and active exchange with all other traffic modeling program researchers in order to facilitate free-flowing information transfer and integration with other investigators. We actively exchanged research results with the larger research community through journal publications and oral presentations at

national/international seminars and meetings.

We documented technical work accomplished and information gained during performance of this acquisition in the form of monthly status and special technical reports and this final report. These reports included significant observations, problems, positive results, negative results, and design criteria. These reports documented procedures followed, processes developed, lessons learned, and other useful information. The reports documented the details of all technical work to permit full understanding of the techniques and procedures used in evolving technology or processes developed.

Task 1: Packet-Level Measurement

We assembled a database of packet-level measurements consisting of packet header fields and time-stamps. The measurements were passive. We obtained existing measurements and deployed devices to collect more data.

We studied best-effort traffic streams, the dominant traffic type on IP networks today, as well as voice over IP traffic. Increasingly, today's voice traffic is moving from the circuit-switched telephone networks to IP networks. Many service providers now have architectures in which calls start and end on the circuit-switched networks, but are carried as IP network packet traffic over the IP network core. Very little measurement of VoIP traffic has been carried out, a prerequisite to characterization and control. We obtained such data as part of our study. For best-effort traffic the headers were IP and TCP (or UDP). For voice over IP traffic, the headers were IP, UDP, and RTP.

Task 2: Statistical Analysis, Data Mining, and Machine Learning for Network Characterization and Modeling

The data from our measurement plans were studied extensively in the S-Net environment. We determined variables that characterize network behavior and developed models for network traffic. The models can provide traffic generation for network simulators.

Task 3: Automated Control

We developed an automated control method for network self-management based on the above variables for network characterization. The performance of the control methods were studied through simulation methods and analytical mathematical methods.

5   Overview: Key Personnel

1. Supported by Grant.

William S. Cleveland, Professor of Statistics and Professor of Computer Science, Purdue University

Michael Zhu, Assistant Professor of Statistics, Purdue University

Hui Chen, Graduate Student, Statistics Department, Purdue University

2. Substantial work was carried out by personnel not supported by the grant.

Bowei Xi, Assistant Professor of Statistics, Purdue University

6 Summary of Technical Results: Data Collection

6.1 VoIP Collection

We deployed a monitor on a 100-megabit VoIP Ethernet link on the Global Crossing Network. The link carries VoIP traffic in two directions between a Sonus gateway and a first router on the GBLX IP network. The hardware is a 100-mbps Ethernet DAG network card that carries out measurement on the card and can measure at full line speed to provide highly accurate time stamps. The DAG card timestamps each packet and captures the IP/UDP/RTP headers. We collected data for 48 hours.

In addition, we obtained call detail records for each call during the first 24 hours of the packet trace. The detail records let us see unequivocally the caller and called directions, the destinations of the calls, and categories of the nature of the call terminations. The call detail records provide information that aid greatly the analysis of the packet traces.

6.2 Best-Effort Collection

We obtained a monitor from the National Laboratory for Applied Network Research (NLANR) for packet trace collection on the 1 gbps link that carries both directions of all traffic between the Purdue University campus and the rest of the Internet. The high-end monitor is a linux box with a monitor operating system together with a light PC in the box that can reboot the monitor operating system and allow increased remote maintenance. The box also contains a 1-gbps DAG card that does timestamping and packet header capture. Joerg Micheel of NLANR flew to the West Lafayette campus to help Scott Ballew and Doug Crabill in the installation. By agreeing to become part of the NLANR measurement program, which means we allow data collection in part for the NLANR database, the box and travel for Micheel was paid for by the NSF grant that supports the NLANR program.

6.3 Database from Previous Collection Efforts

In addition, we built a database of best effort packet traces from 6 Internet links. This live stream database consists of 349 streams, 90 sec or 5 min in duration, from 6 links that we name BELL, NZIX, AIX1, AIX2, MFN1, and MFN2. The measured streams have negligible delay on the link input router. The mean number of simultaneous active connections, c, ranges from 49 connections to 18,976 connections. The traffic bit rate, τ, ranges from 1.00 megabits/sec to 348 megabits/sec.

BELL is a 100 megabit/sec link in Murray Hill, NJ, USA connecting a Bell Labs local network of about 3000 hosts to the rest of the Internet. The transmission is half-duplex, so both directions, in and out, are multiplexed and carried on the same link, and a stream is the multiplexing of both directions; but to keep the variable c commensurate for all of the six links, the two directions for each connection is counted as two. In this presentation we use 195 BELL traces, each 5 min in length. NZIX is the 100 megabit/sec New Zealand Internet exchange hosted by the ITS department at the University of Waikato, Hamilton, New Zealand, that served as a peering point among a number of major New Zealand Internet Service Providers at the time of data collection. All arriving packets from the input/output ports on the switch are mirrored, multiplexed, and sent to a port where they are measured. Because all connections have two directions at the exchange, each connection counts 2 as for BELL. In this presentation we use 84 NZIX traces, each 5 min in

5

length. AIX1 and AIX2 are two separate 622 megabit/sec OC12 packet-over-sonet links, each carrying one direction of the traffic between NASA Ames and the MAE-West Internet exchange. In this presentation we use 23 AIX1 and 23 AIX2 traces, each 90 sec in length. AIX1 and AIX2 streams were collected as part of a project at the National Laboratory for Applied Network Research where the data are collected in blocks of 90 sec. MFN1 and MFN2 are two separate 2.5 gigabit/sec OC48 packet-over-sonet links on the network of the service provider Metropolitan Fiber Network; each link carries one direction of traffic between San Jose, California, USA and Seattle, Washington, USA.

We also downloaded from NLANR data collected on the OC3 link that connects the University of Leipzig network to the rest of the Internet.

7 Summary of Technical Results: Statistical Analysis, Data Mining, and Machine Learning for Network Characterization and Modeling

7.1 Algorithms for Data Processing and Strategies of Analysis

We purchased and installed a high-end two-processor linux box for analysis, learning, mining, and modeling of the trace data. We installed the S-Net system for graphics and data analysis on the box.

Algorithms were written in C, Perl, and the S language for graphics and data analysis to process the VoIP measurements. The algorithms represent a a dataflow that progresses from raw packet trace files, to primary and secondary UNIX flatfile databases, and then to S objects. Our analysis is carried out within the S environment to allow comprehensive study.

An important part of the strategy for study is time blocking. We broke the trace and CDR data into time blocks, and applied the same analysis to all blocks. We used 5 minute blocks to guarantee stationarity of traffic processes within each interval and to create object sizes that were not to large for efficient computation.

7.2 Analysis for VoIP

We used tools of statistical analysis, data mining, and machine learning to study the characteristics of the VoIP packet trace data. Most important in this exploratory, or first-study phase, were a host of visualization tools.

We discovered that there were a very large number of packet level calls with no CDR records. They are all very short. We studied their characteristics to see if they had anomalous behavior. They spread out across time and space in proportion to the calls with CDR records. We concluded that certain calls short enough not to generate revenue are not saved in the CDR records.

The study consisted of a wide range of topics involving the VoIP traffic characteristics and the quality performance of the Global Crossing system

- call arrival process, intensity and detailed statistical behavior
- call characteristics such as numbers of attempts (without connection) and connections, call durations, and detailed behavior of packet-level characteristics during ringing period vs. during talking period

6

- packet jitter
- aggregate packet traffic statistical properties of inter-arrival times and packet counts in 20 ms intervals
- silence suppression

A number of results came from our analysis.

- call inter-arrivals have to a very good approximation an exponential marginal distribution
- the call inter-arrivals have a small amount of short-term positive autocorrelation
- the call durations have a Pareto distribution, which breaks with the past exponential behavior, caused because today's call contain faxes and computer logins, not just the voice conversations of the past
- a very large fraction of calls are call attempts
- patterns in the ringing make us believe we can develop a classification algorithm for determining the ringing period
- the level of jitter is very low, about $\pm 150 \mu$sec.
- the packet arrival process is long-range dependent, caused by the Pareto marginal distribution of the call durations; these statistical properties play an important role in the automated control rules for bandwidth allocation that we are developing.

7.3 FSD Models for Best-Effort Traffic

The statistical properties of best-effort traffic, the packet arrival process, with its long-range dependence, has not reached a consensus with different camps arguing for different views.

Almost all analysis in the past has consisted of the study of counts of packet arrivals in a fixed interval of time such as 1ms or 10 ms. This is not sufficient to understand the arrival process because moving to a count is an information reduction method to simplify analysis. Instead, we studied and modeled all of the information, the packet arrival times.

We developed a new class of models, fractional sum-difference (FSD) models, for the inter-arrival times. The FSD models, which are non-Gaussian and long-range dependent, provide excellent fits to the inter-arrivals. We carried out extensive validation to demonstrate this. Our models provide a bridge between the different camps, explaining and integrating results on multifractal modeling, long-range dependence, fractional Brownian motion, and convergence to Poisson processes.

In addition, our models can be used for open-loop generation of background traffic in network simulators. This represents the first such generation that properly recreates the properties of real traffic.

Suppose x_v for $v = 1, 2, \ldots$ is a stationary time series with marginal cumulative distribution function $F(x; \phi)$ where ϕ is a vector of unknown parameters. Let $x_v^* = H(x_v; \phi)$ be a transformation of x_v such that the marginal distribution of x_v^* is normal with mean 0 and variance 1. We have $H(x_v; \phi) = G^{-1}(F(x; \phi))$, where $G(z)$ is the cumulative distribution function of a normal random variable with mean 0 and variance 1. Next we suppose x_v^* is a Gaussian time series and call x_v^* the "Gaussian image" of x_v.

Suppose x_v^* has the following form:

$$x_v^* = \sqrt{1 - \theta}\, s_v + \sqrt{\theta}\, n_v,$$

where s_v and n_v are independent of one another and each has mean 0 and variance 1. n_v is Gaussian white noise, that is, an independent time series. s_v is a Gaussian fractional ARIMA (Hosking 1981)

$$(I - B)^d s_v = \epsilon_v + \epsilon_{u-1}$$

where $Bs_v = s_{u-1}$, $0 < d < 0.5$, and ϵ_v is Gaussian white noise with mean 0 and variance

$$\sigma_\epsilon^2 = \frac{(1 - d)\Gamma^2(1 - d)}{2\Gamma(1 - 2d)}.$$

x_v is a *fractional sum-difference* (*FSD*) time series. Its Gaussian image, x_v^*, has two components. $\sqrt{1 - \theta}s_v$ is the long-range-dependent, or lrd, component; its variance is $1 - \theta$. $\sqrt{\theta}$ is the white-noise component; its variance is θ.

Let $p_{x^*}(f)$ be the power spectrum of the x_v^*. Then

$$p_{x^*}(f) = (1 - \theta)\sigma_\epsilon^2 \frac{4\cos^2(\pi f)}{(4\sin^2(\pi f))^d} + \theta$$

for $0 \le f \le 0.5$. As $f \to 0.5$, $p_{x^*}(f)$ decreases monotonically to θ. As $f \to 0$, $p_{x^*}(f)$ goes to infinity like $\sin^{-2d}(\pi f) \sim f^{-2d}$, one outcome of long-range dependence. For nonnegative integer lags k, let $r_{x^*}(k)$, $r_s(k)$, and $r_n(k)$ be the autocovariance functions of x_v^*, s_v, and n_v, respectively. Because the three series have variance 1, the autocovariance functions are also the autocorrelation functions. $r_s(k)$ is positive and falls off like k^{2d-1} as k increases, another outcome of long-range dependence. For $k > 0$, $r_n(k) = 0$ and

$$r_{x^*}(k) = (1 - \theta)r_s(k).$$

As $\theta \to 1$, x_v^* goes to white noise: $p_{x^*}(f) \to 1$ and $r_{x^*}(k) \to 0$ for $k > 0$. The change in the autocovariance function and power spectrum are instructive. As θ gets closer to 1, the rise of $p_{x^*}(f)$ near $f = 0$ is always to order f^{-2d}, and the rate of decay of $r_{x^*}(k)$ for large k is always k^{2d-1}; but the ascent of $p_{x^*}(f)$ at the origin begins closer and closer to $f = 0$ and the $r_{x^*}(k)$ get uniformly smaller by the multiplicative factor $1 - \theta$.

We model the marginal distribution of t_v by a Weibull with shape λ and scale α, a family with two unknown parameters. Estimates of λ are almost always less than 1. The Weibull provides an excellent approximation of the sample marginal distribution of the t_v except that the smallest 3% to 5% of the sample distribution is truncated to a nearly constant value due to certain network transmission properties.

The marginal distribution of q_v is modeled as follows. While packets less than 40 bytes can occur, it is sufficiently rare that we ignore this and suppose $40 \le q_v \le 1500$. First, we provide for A atoms at sizes $\phi_1^{(s)} \ldots \phi_A^{(s)}$ such as 40 bytes, 512 bytes, 576 bytes, and 1500 bytes that are commonly occurring sizes; the atom probabilities are $\phi_1^{(a)} \ldots \phi_A^{(a)}$. For the remaining sizes, we divided the interval [40, 1500] bytes into C sub-intervals using $C - 1$ distinct break points with values that are greater than 40 bytes and less than 1500 bytes, $\phi_1^{(b)}, \ldots \phi_{C-1}^{(b)}$. For each of the C sub-intervals, the size distribution has an equal probability for the remaining sizes (excluding the atoms) in the sub-interval; the total probabilities for the sub-intervals are $\phi_1^{(i)}, \ldots, \phi_C^{(i)}$. Typically, with just 3 atoms at 40 bytes, 576 bytes, and 1500 bytes, and with just 2 break points at 50 bytes and 200 bytes, we get an excellent approximation of the marginal distribution.

The transformed time series t_v^* and q_v^* appear to be quite close to Gaussian processes. Some small amount of non-Gaussian behavior is still present but it is minor. The autocorrelation structure of these Gaussian images are very well fitted by the FSD autocorrelation structure.

The parameters of the FSD model are the following:

- q_v marginal distribution: A atom probabilities $\phi_i^{(a)}$ at A sizes $\phi_i^{(s)}$; $C - 1$ break points $\phi_i^{(b)}$ and C sub-interval probabilities $\phi_i^{(i)}$
- t_v marginal distribution: shape λ and scale α
- q_v^* time dependence: fractional difference coefficient $d^{(q)}$ and white-noise variance $\theta^{(q)}$
- t_v^* time dependence: fractional difference coefficient $d^{(t)}$ and white-noise variance $\theta^{(t)}$.

We found that the $d^{(q)}$ and $d^{(t)}$ do not depend on c; this is based on empirical study, and it is supported by theory. The estimated values are 0.410 and 0.411, respectively. We take the value of each of these two parameters to be 0.41. We found that as c increases, estimates of λ, $\theta^{(q)}$, and $\theta^{(t)}$ all tend toward 1. This means the t_v tend to a independent exponentials, a Poisson process and the q_v tend toward independence. In other words, the statistical models account for the change in the t_v and q_v and c increases that was discussed earlier. We estimated these three parameters and α by partial likelihood methods with $d^{(q)}$ and $d^{(t)}$ fixed to 0.41. The marginal distribution of q_v on a given link does not change with c, but it does change from link to link. To generate traffic, we must specify the atom and sub-interval probabilities. This provides a mean packet size $\mu_{(q)}$, measured in bits/packet.

8 Automated Control: Bandwidth Estimation for Best-Effort Traffic

8.1 Problem Formulation

Suppose the packet stream — packet arrival times and sizes — arriving for transmission on an Internet link is best-effort traffic with bit rate τ bits/sec and number of simultaneous active connections, c. Suppose the link input buffer is large enough that packet loss is negligible. Our goal is to estimate the quality-of-service (QoS) bandwidth β in bits/sec, or equivalently, the QoS utilization $u = \tau/\beta$, that satisfies QoS criteria for the packet queueing delay in the link input buffer. The criteria are a delay, δ sec, and the probability, ω, that the delay for a packet exceeds δ.

8.2 Other Work on the Problem

There is a wide literature on the bandwidth estimation problem. Much of it is theoretical — mathematical results that derive properties of queueing systems. A smaller literature is empirical in nature, based on simulations with packet stream inputs from measurements on live links or from models for traffic. The classical Erlang Delay Formula provides a simple formula that can be used for estimation of traffic streams that in theory have Poisson arrivals and i.i.d. exponential sizes. Best-effort traffic is much more complex, it is nonlinear and long-range dependent, and there has been to date no simple, validated formula for it.

8.3 Principal Result: The Best-Effort Delay Model

The principal result of this report is a statistical model that provides a simple, validated formula for the estimation of bandwidth for best-effort traffic in the same way that the Erlang Delay Formula does for the

Poisson-exponential case. The model has been validated through extensive empirical study and through consistency with certain theoretical properties of queueing.

The model consists of the Best-Effort Delay Formula plus random variation:

$$\text{logit}_2(u) = o + o_c \log_2(c) + o_{\tau\delta} \log_2(\tau\delta) + o_\omega(-\log_2(-\log_2(\omega))) + \psi.$$

ψ is a random error variable with mean 0 and median absolute deviation $m_\delta(\psi)$, which depends on δ; $\log_2$ is the log base 2; and $\text{logit}_2(u) = \log_2(u/(1-u))$. The distribution of $0.691\psi/m_\delta(\psi)$ is a t-distribution with 15 degrees of freedom. Estimates of the coefficients of the model are

$$\hat{o} = -8.933 \quad \hat{o}_c = 0.420 \quad \hat{o}_{\tau\delta} = 0.444 \quad \hat{o}_\omega = 0.893.$$

$m_\delta(\psi)$ is modeled as a function of δ: $\log_2(m_\delta(\psi))$ is a linear function of $\log_2(\delta)$ plus random variation. The estimate of the intercept of the line is –0.481, the estimate of the linear coefficient of the line is 0.166, and the estimate of the standard error is 0.189. The bit rate τ is equal to $c\gamma_b$ where γ_b is the connection bit rate in bits/sec/connection. So the Best-Effort Delay Formula can also be written as

$$\text{logit}_2(u) = o + (o_c + o_{\tau\delta})\log_2(c) + o_{\tau\delta}\log_2(\gamma_b\delta) + o_\omega(-\log_2(-\log_2(\omega))).$$

In this form we see the action of the amount of multiplexing of connections as measured by c, and the end-to-end connection speed as measured by γ_b. An increase in either results in an increase in the utilization of a link.

The Best-Effort Delay Model is used to estimate the bandwidth required to carry best-effort traffic given δ, ω, τ, and c. The QoS logit utilization is estimated by

$$\hat{\ell} = -8.933 + 0.420\log_2(c) + .444\log_2(\tau\delta) + 0.893(-\log_2(-\log_2(\omega))),$$

so the QoS utilization is estimated by

$$\hat{u} = \frac{2^{\hat{\ell}}}{1 + 2^{\hat{\ell}}}.$$

The corresponding estimated bandwidth is $\tau/\hat{u}$. For such an estimate there is a 50% chance of being too large and a 50% chance of being too small. We might however, use a more conservative estimate that provides a much smaller chance of too little bandwidth. Let

$$\hat{m}_\delta(\psi) = 2^{-0.481 + 0.166\log_2(\delta)}$$

be the estimate of $m(\delta)$. Let $t_{15}(p)$ be the lower $100p\%$ percentage point of a t-distribution with 15 degrees of freedom, where p is small, say 0.05. Let

$$\hat{\ell}(p) = \hat{\ell} - \hat{m}(\delta)t_{15}(p)/0.691$$

Then

$$\hat{u}(p) = \frac{2^{\hat{\ell}(p)}}{1 + 2^{\hat{\ell}(p)}},$$

is a conservative utilization estimate, the lower limit of a $100p\%$ tolerance interval for the QoS utilization. The corresponding estimated bandwidth is $\tau/\hat{u}(p)$.

8.4   Methods

The Best-Effort Delay Model was built, in part, from queueing theory. Certain predictor variables were suggested by the Erlang Delay Formula. Theory prescribes certain behavior as τ, c, or γ_b increase, resulting in rate gains, multiplexing gains, or fast-forward invariance, and the model was constructed to reproduce the behavior.

The Best-Effort Delay Model was built, in part, from results of queueing simulations with traffic stream inputs of two types, live and synthetic. The live streams are measurements of packet arrivals and sizes for 349 intervals, 90 sec or 5 min in duration, from 6 Internet links. The synthetic streams are arrivals and sizes generated by recently developed FSD time series models for the arrivals and sizes of best-effort traffic. Each of the live streams was fitted by two FSD models, one for the inter-arrivals and one for the sizes, and a synthetic stream of 5 min was generated by the models. The generated inter-arrivals are independent of the generated sizes, which is what we found in the live data. The result is 349 synthetic streams that match the statistical properties collectively of the live streams. For each live or synthetic stream, we carried out 25 runs, each with a number of simulations. For each run we picked a delay δ and a delay probability ω; simulations were carried out to find the QoS bandwidth β, the bandwidth that results in delay probability ω for δ. This also yields a QoS utilization $u = \tau/\beta$. We used 5 delays — 0.001 sec, 0.005 sec, 0.010 sec, 0.050 sec, 0.100 sec — and 5 delay probabilities — 0.001, 0.005, 0.01, 0.02, and 0.05 — employing all 25 combinations of the two delay criteria. The queueing simulation results in delay data: values of five variables — QoS utilization u, delay δ, delay probability ω, the mean number of active connections of the traffic c, and the traffic bit rate τ. The delay data were used in the model building.

8.5   Validity and Applicability

Extensive data exploration with visualization tools, some shown here, demonstrate that the Best-Effort Delay Model fits the simulation delay data. This, of course, is necessary for the model to be valid. In addition, validity is supported by the model reproducing the theoretical queueing properties as just discussed.

The validity of the Best-Effort Delay Model depends on the validity of the traffic streams used as inputs to the queueing simulation; that is, the packet streams must reproduce the statistical properties of best-effort streams. Of course, the live streams of the study do so because they are best-effort traffic. Extensive validation has shown that the FSD models, used to generate the packet streams here, provide excellent fits to best-effort packet streams when c is above about 64 connections, which for a link where γ_b is about 2^{14} bits/sec/connection means τ is above about 1 megabit/sec. For this reason, only traffic streams with τ greater than this rate are used in the study, and the Best-Effort Delay Model is valid above this rate.

The results are only valid for links with a large enough buffer that the packet loss is negligible. We have used open-loop study, which does not provide for the TCP feedback that occurs when loss is significant.

There is also a practical restriction on applicability. We have taken the range of our study to include traffic bit rates as low as about 1 megabit/sec. We have done this simply because we can do so and achieve valid results. But even for the least stringent of our delay criteria, $\delta = 0.1$ sec delay and $\omega = 0.05$ delay probability, the utilizations are low for rates in the range of 1 megabit/sec to 5 megabits/sec. This might well be judged to be too small a utilization to be practical; if so, it might mean that the what must be sacrificed is the negligible packet loss, which means that a QoS study at the very low traffic bit rates needs to take account of TCP feedback.

One outcome of the dependence of the bandwidth estimation on the traffic statistics is that our solution for best-effort traffic would not apply to other forms of Internet traffic that do not share the best-effort statistical properties. One example is voice traffic.

Finally, the Best-Effort Delay Model provides an estimation of bandwidth in isolation without considering other network factors. A major factor in network design is link failures. Redundancy needs to be built into the system. So an estimate of bandwidth from the model for a link based on the normal link traffic may be reduced to provide the redundancy. But the model still plays a role because the bandwidth must be chosen based on link traffic, but now the traffic in the event of a failure elsewhere.

9 Automated Control: Bandwidth Estimation for VoIP

Today, no automated methods exist for determining how many calls can be put on a VoIP link and maintain QoS criteria for delay and packet jitter. We designed a queueing simulation and modeling mechanism to determine a statistical process-control approach that would allow automation provisioning of traffic on a link.

In the simulation, calls arrive as a Poisson process, a reasonable approximation to the observed behavior of the arrivals process. For each arrival, we sample with replacement from our call database, about 300,000 calls, and use the observed packet process of each sampled call in the simulation.

In the simulation we fixed the traffic call rate and vary the link bandwidth (link speed) so that the utilization ranges from 30% to 99%, and study the resulting queueing delay and jitter of the packets. We built statistical models for the delay and jitter based on the utilization, the number of simultaneous active calls and the bitrate. The models provide the requisite statistical process control.

We studied our simulation runs and discovered that there are instabilities when the utilization gets beyond a certain limit, which makes delay estimation problematic. The limit decreases and the traffic rate decreases. The phenomenon means the network cannot operate at these utilizations, so they are not relevant. We cannot leave them in the data to be modeled because they depart from standard conditions, so we will eliminated them from the study, using machine learning rules for the elimination.

10 Work Under The Current Contract

This report describes the work done under the above federal contract as well as previous work done under federal contract No. F30602-02-2-C0093 awarded to Bell Labs because the work under the new contract was a continuation of the work under the previous contract.

The following describes that portion of the work completed under the new contract:
Task 1: Packet-Level Measurement

- Newark VoIP packet trace data collection and data processing
- Leipzig best-effort traffic processing
- installation of NLANR packet trace collector on Purdue Internet access link for best-effort traffic measurement

12

Task 2: Statistical Analysis, Data Mining, and Machine Learning for Network Characterization and Modeling

- all analysis of the Newark VoIP data
- all work on the integration of best-effort traffic models through the understanding provided by the fractional sum-difference models

Task 3: Automated Control

- all simulations for bandwidth estimation for VoIP traffic and ensuing analysis and modeling of results
- simulations for best-effort traffic were completely redesigned and carried out, and new results analyzed and modeled

11 Detailed Technical Report

11.1 Introduction: Contents of the Paper

The Internet is a world-wide computer network; at any given moment, a vast number of pairs of hosts are transferring files one to the other. Each transferred file is broken up into packets that are sent along a path across the Internet consisting of links and nodes. The first node is the sending host; a packet exits the host and travels along a link (fiber, wire, cable, or air) to a first router node, then over a link to a second router node, and so forth until a last router sends the packet to a receiving host node over a final link.

The packet traffic arriving for transmission on an Internet link is a stream: a sequence of packets with arrival times (sec) and sizes (bytes or bits). The packets come from pairs of hosts using the link for their transfers; that is, the link lies on the path from one host to another for each of a collection of pairs of hosts. When a packet arrives for transmission on a link, it enters a buffer (bits) where it must wait if there are other packets waiting for transmission or if a packet is in service, that is, in the process of moving out of the buffer onto the link. If the buffer is full, the packet is dropped.

A link has a bandwidth (bits/sec), the rate at which the bits of a packet are put on the link. Over an interval of time during which the traffic is stationary, the packets arrive for transmission at a certain rate, the traffic bit rate (bits/sec), which is defined, formally, to be the mean of the packets sizes (bits) divided by the mean packet inter-arrival time (sec), but this is approximately the mean number of arriving bits over the interval divided by the interval length (sec). Over the interval there is a mean simultaneous active connection load, the mean number of source-destination pairs of hosts actively sending packets over the link. The utilization of the link is the traffic bit rate divided by the bandwidth; it measures the traffic rate relative to the capacity of the link.

This article presents results on a fundamental problem of engineering the Internet. What link bandwidth is needed to accommodate traffic with a certain bit rate and ensure that the transmission on the link maintains quality-of-service, or QoS, criteria? Finding the QoS bandwidth must be faced in setting up every link on the Internet, from the low-bandwidth links connecting the computers of a home user to the high-bandwidth links of a major Internet service provider. Our approach to solving the bandwidth estimation problem is to use queueing theory and queueing simulations to build a model for the QoS bandwidth. The traffic inputs are live streams from measurements of live links and synthetic streams from statistical models for traffic streams.

13

Section 11.2 describes Internet TCP/IP transmission technology, which governs almost all computer networking today — for example, the networks of Internet service providers, universities, companies, and homes. Section 11.2 also describes the buffer queueing process and its affect on the QoS of file transfer.

Section 11.3 formulates the particular version of the bandwidth estimation problem that is addressed here, discusses why the statistical properties of the packet streams are so critical to bandwidth estimation, and outlines how we use queueing simulations to study the problem. We study "best-effort" Internet traffic streams because they are the predominant type of traffic on Internet links today. The QoS criteria for best-effort streams are the packet loss (fraction of lost packets), the queueing delay (sec), and the delay probability (probability of a packet exceeding the delay). We suppose that the link packet loss is negligible, and find the QoS bandwidth required for a packet stream of a certain load that satisfies the delay and the delay probability.

Section 11.4 describes fractional sum-difference (FSD) time series models, which are used to generate the synthetic streams for the queueing simulations. The FSD models, a new class of non-Gaussian, long-range dependent time series models, provide excellent fits to packet size time series and to packet inter-arrival time series. The validation of the FSD models is critical to this study. The validity of our solution to the bandwidth estimation problem depends on having traffic inputs to the queueing that reproduce the statistical properties of best-effort traffic. Of course, the live data have the properties, but we need assurance that the synthetic data do as well.

Section 11.5 describes the live packet arrivals and sizes and the synthetic packet arrivals and sizes that are generated by the FSD models. Section 11.6 gives the details of the simulations and the resulting delay data: values of the QoS bandwidth, delay, delay probability, mean number of active host-pair connections of the traffic, and traffic bit rate.

Model building, based on the simulation delay data and on queueing theory, begins in Section 11.7. To do the model building and diagnostics, we exploit the structure of the delay data — utilizations for all combinations of delay and delay probability for each stream, live or synthetic. We develop an initial model that relates, for each stream, the QoS utilization (bit rate divided by the QoS bandwidth) to the delay and delay probability. We find a transformation for the utilization for which the functional dependence on the delay and delay probability does not change with the stream. There is also an additive stream coefficient that varies across streams, characterizing the statistical properties of each stream. This stream-coefficient delay model cannot be used for bandwidth estimation because the stream coefficient would not be known in practice.

Next we add two variables to the model that measure the statistical properties of the streams and that can be specified or measured in practice — the traffic bit rate and the number of simultaneous active host-pair connections on the link — and drop the stream coefficients. In effect we have modeled the coefficients. The result is the Best-Effort Delay Model: a Best-Effort Delay Formula for the utilization as a function of (1) the delay, (2) the delay probability, (3) the traffic bit rate, and (4) the mean number of active host-pair connections of the traffic, plus a random error term.

Section 11.8 presents a method for bandwidth estimation that starts with the value from the Best-Effort Delay Formula, and then uses the error distribution of the Best-Effort Delay Model to find a tolerance interval whose minimum value provides an conservative estimate with a low probability of being too small.

Section 11.9 discusses previous work on bandwidth estimation, and how it differs from the work here. Section 11.10 is an extended abstract. Readers who seek just results can proceed to this section; those not familiar with Internet engineering technology might want to first read Sections 11.2 and 11.3.

14

The following notation is used throughout the article:

Packet Stream

- v : arrival numbers (number) $v = 1$ is the first packet, $v = 2$ is the second packet, etc.
- a_v : arrival times (sec)
- t_v : inter-arrival times (sec) $t_v = a_{v+1} - a_v$
- q_v : sizes (bytes or bits).

Traffic Load

- c : mean number of simultaneous active connections (number)
- τ : traffic bit rate (bits/sec)
- γ_p : connection packet rate (packets/sec/connection)
- γ_b : connection bit rate (bits/sec/connection).

Bandwidth

- β : bandwidth (bits/sec)
- u : utilization (fraction) τ/β.

Queueing

- δ : packet delay (sec)
- ω : delay probability (fraction).

11.2   Internet Technology

The Internet is a computer network over which a pair of host computers can transfer one or more files (Stevens 1994). Consider the downloading of a Web page, which is often made up of more than one file. One host, the client, sends a request file to start the downloading of the page. Another host, the server, receives the request file and sends back a first response file; this process continues until all of the response files necessary to display the page are sent. The client passes the received response files to a browser such as Netscape, which then displays the page on the screen. This section gives information about some of the Internet engineering protocols involved in such file transfer.

11.2.1   Packet Communications

When a file is sent, it is broken up into packets whose sizes are 1460 bytes or less. The packets are sent from the source host to the destination host where they are reassembled to form the original file. They travel along a path across the Internet that consists of transmission links and routers. The source computer is connected to a first router by a transmission link; the first router is connected to a second router by another transmission link, and so forth. A router has input links and output links. When it receives a packet from one of its input links, it reads the destination address on the packet, determines which of the routers connected to it by output links should get the packet, and sends out the packet over the output link connected to that

router. The flight across the Internet ends when a final router receives the packet on one of its input links and sends the packet to the destination computer over one of its output links.

The two hosts establish a connection to carry out one or more file transfers. The connection consists of software running on the two computers that manages the sending and receiving of packets. The software executes an Internet transport protocol, a detailed prescription for how the sending and receiving should work. The two major transport protocols are UDP and TCP. UDP just sends the packets out. With TCP, the two hosts exchange control packets that manage the connection. TCP opens the connection, closes it, retransmits packets not received by the destination, and controls the rate at which packets are sent based on the amount of retransmission that occurs. The transport software adds a header to each packet that contains information about the file transfer. The header is 20 bytes for TCP and 8 bytes for UDP.

Software running on the two hosts implements another network protocol, IP (Internet Protocol), that manages the involvement of the two hosts in the routing of a packet across the Internet. The software adds a 20-byte IP header to the packet with information needed in the routing such as the source host IP address and the destination host IP address. IP epitomizes the conceptual framework that underlies Internet packet transmission technology. The networks that make up the Internet — for example, the networks of Internet service providers, universities, companies, and homes — are often referred to as IP networks, although today it is unnecessary because almost all computer networking is IP, a public-domain technology that defeated all other contenders, including the proprietary systems of big computer and communications companies.

11.2.2 Link Bandwidth

The links along the path between the source and destination hosts each have a bandwidth β in bits/sec. The bandwidth refers to the speed at which the bits of a packet are put on the link by a computer or router. For the link connecting a home computer to a first router, β might be 56 kilobits/sec if the computer uses an internal modem, or 1.5 megabits/sec if there is a broadband connection, a cable or DSL link. The link connecting a university computer to a first router might be 10 megabits/sec, 100 megabits/sec, or 1 gigabit/sec. The links on the core network of a major Internet service provider have a wide range of bandwidths; typical values range from 45 megabits/sec to 10 gigabits/sec. For a 40 byte packet, which is 320 bits, it takes 5.714 ms to put the packet on a 56 kilobit/sec link, and 0.032 μs to put it on a 10 gigabits/sec link, which is about 180,000 times faster. Once a bit is put on the link, it travels down the link at the speed of light.

11.2.3 Active Connections, Statistical Multiplexing, and Measures of Traffic Loads

At any given moment, an Internet link has a number of simultaneous active connections; this is the number of pairs of computers connected with one another that are sending packets over the link. The packets of the different connections are intermingled on the link; for example, if there are three active connections, the arrival order of 10 consecutive packets by connection number might be 1, 1, 2, 3, 1, 1, 3, 3, 2, and 3. The intermingling is referred to as "statistical multiplexing". On a link that connects a local network with about 500 users there might be 300 active connections during a peak period. On the core link of an Internet service provider there might be 60,000 active connections.

During an interval of time when the traffic is stationary, there is a mean number of active connections c, and a traffic bit rate τ in bits/sec. Let $\mu_{(t)}$ in sec be the mean packet inter-arrival time, and let $\mu_{(q)}$, in bits, be the mean packet size. Then the packet arrival rate per connection is $\gamma_p = c^{-1}\mu_{(t)}^{-1}$ packets/sec/connection.

The bit rate per connection is $\gamma_b = \mu_{(q)}c^{-1}\mu_{(t)}^{-1} = \tau c^{-1}$ bits/sec/connection. γ_p and γ_b measure the average host-to-host speed of Internet connections — for example, the rate at which the file of a page is downloaded — for the pairs of hosts that use the link.

The bit rate of all traffic on the link is $\tau = c\gamma_b$. Of course, $\tau \leq \beta$ because bits cannot be put on the link at a rate faster than the bandwidth. A larger traffic bit rate τ requires a larger bandwidth β. Let us return to the path across the Internet for the Web page download discussed earlier. Starting from the link connecting the client computer to the Internet and proceeding though the links, τ tends to increase, and therefore, so does β. We start with a low bandwidth link, say 1.5 megabits/sec, then to a link at the edge of a service provider network, say 156 megabits/sec, and then to the core links of the provider, say 10 gigabits/sec. As we continue further, we move from the core to the service provider edge to a link connected to the destination computer, so τ and β tend to decrease.

11.2.4 Queueing, Best-Effort Traffic, and QoS

A packet arriving for transmission on a link is presented with a queueing mechanism. The service time for a packet is the time it takes to put the packet on the link, which is the packet size divided by the bandwidth β. If there are any packets whose transmission is not completed, then the packet must wait until these packets are fully transmitted before its transmission can begin. This is the queueing delay. The packets waiting for transmission are stored in a buffer, a region in the memory of the computer or router. The buffer has a size. If a packet arrives and the buffer is full, then the packet is dropped. As we will see, the arrival process for packets on a link is long-range dependent; at low loads, the traffic is very bursty, but as the load increases, the burstiness dissipates. For a fixed τ and β, bursty traffic results in a much larger queue-height distribution than traffic with Poisson arrivals.

TCP, the predominant protocol for managing file transfers, changes the rate at which it sends packets with file contents. TCP increases the rate when all goes well, but reduces the rate when a destination computer indicates that a packet has not been received; the assumption is that congestion somewhere on the path has led to a buffer overflow, and the rate reduction is needed to help relieve the congestion. In other words, TCP is closed-loop because there is feedback. UDP is not aware of dropped packets and does not respond to them.

When traffic is sent across the Internet using TCP or UDP and this queueing mechanism, with no attempts to add additional protocol features to improve QoS, then the traffic is referred to as "best-effort". IP networks are a best-effort system because the standard protocols make an effort to get packets to their destination, but packets can be delayed, lost, or delivered out of order. Queueing delay and packet drops degrade the QoS of best-effort traffic. For example, for Web page transfers, the result is a longer wait by the user, partly because the packets sit in the queue, and partly because of TCP reducing its sending rate when retransmission occurs. Best-effort traffic contrasts with priority traffic, which when it arrives at a router, goes in front of best-effort packets. Packets for voice traffic over the Internet are often given priority.

11.3 The Bandwidth Estimation Problem: Formulation and Stream Statistical Properties

11.3.1 Formulation

Poor QoS resulting from delays and drops on an Internet link can be improved by increasing the link bandwidth β. The service time decreases, so if the traffic rate τ remains fixed, the queuing delay distribution decreases, and delay and loss are reduced. Loss and delay are also affected by the buffer size; the larger the buffer size, the fewer the drops, but then the queueing delay has the potential to increase because the maximum queueing delay is the buffer size divided by β.

The bandwidth estimation problem is to choose β to satisfy QoS criteria. The resulting value of β is the QoS bandwidth. The QoS utilization is the value of $u = \tau/\beta$ corresponding to the QoS bandwidth. When a local network, such as a company or university, purchases bandwidth from an Internet service provider, a decision on β must be made. When an Internet service provider designs its network, it must choose β for each of its links. The decision must be based on the traffic load and QoS criteria.

Here, we address the bandwidth estimation problem specifically for links with best-effort traffic. We take the QoS criteria to be delay and loss. For delay, we use two metrics: a delay, δ, and the delay probability, ω, the probability that a packet exceeds the delay. For loss, we suppose that the decision has been made to choose a buffer size large enough that drops will be negligible. This is, for example, consistent with the current practice of service providers on their core links (Iyer, Bhattacharyya, Taft, McKeown, and Diot 2003). Of course, a large buffer size allows the possibility of a large delay, but setting QoS values for δ and ω allows us to control delay probabilistically; we will chose ω to be small. The alternative is to use the buffer size as a hard limit on delay, but because dropped packets are an extreme remedy with more serious degradations of QoS, it is preferable to separate loss and delay control, using the softer probabilistic control for delay. Stipulating that packet loss will be negligible on the link means that for a connection that uses the link, another link will be the loss bottleneck; that is, if packets of the connection are dropped, it will be on another link. It also means that TCP feedback can be ignored in studying the bandwidth estimation problem.

11.3.2 Packet Stream Statistical Properties

A packet stream consists of a sequence of arriving packets, each with a size. Let v be the arrival number; v =1 is the first packet, $v = 2$ is the second packet, and so forth. Let a_v be the arrival times, let $t_v = a_{v+1} - a_v$ be the inter-arrival times, and let q_v be the size of the packet arriving at time a_v. The statistical properties of the packet stream will be described by the statistical properties of t_v and q_v as time series in v.

The QoS bandwidth for a packet stream depends critically on the statistical properties of the t_v and q_v. Directly, the bandwidth depends on the queue-length time process. But the queue-length time process depends critically on the stream statistical properties. Here, we consider best-effort traffic. It has persistent, long-range dependent t_v and q_v (Ribeiro, Riedi, Crouse, and Baraniuk 1999; Gao and Rubin 2001; Cao, Cleveland, Lin, and Sun 2001). Persistent, long-range dependent t_v and q_v have dramatically larger queue-size distributions than those for independent t_v and q_v (Konstantopoulos and Lin 1996; Erramilli, Narayan, and Willinger 1996; Cao, Cleveland, Lin, and Sun 2001). The long-range dependent traffic is burstier than the independent traffic, so the QoS utilization is smaller because more headroom is needed to allow for the bursts. This finding demonstrates quite clearly the impact of the statistical properties. But a corollary of the finding is that the results here are limited to best-effort traffic streams (or any other streams with similar statistical properties). Results for other types of traffic with quite different statistical properties —

for example, links carrying voice traffic using current Internet protocols — are different.

Best-effort traffic is not homogeneous. As the traffic connection load c increases, the arrivals tend toward Poisson and the sizes tend toward independent(Cao, Cleveland, Lin, and Sun 2002; Cao and Ramanan 2002). The reason is the increased statistical multiplexing of packets from different connections; the intermingling of the packets of different connections is a randomization process that breaks down the correlation of the streams. In other words, the long-range dependence dissipates. This means that in our bandwidth estimation study, we can expect a changing estimation mechanism as c increases. In particular, we expect "multiplexing gains", greater utilization due to the reduction in dependence. Because of the change in properties with c, we must be sure to study streams with a wide range of values of c.

11.4   FSD Time Series Models for Packet Arrivals and Sizes

This section presents FSD time series models, a new class of non-Gaussian, long-range dependent models (Cao, Cleveland, Lin, and Sun 2002; Cao, Cleveland, and Sun 2004). The two independent packet-stream time series — the inter-arrivals, t_v, and the sizes, q_v — are each modeled by an FSD model; the models are used to generate synthetic best-effort traffic streams for the queueing simulations of our study.

There are a number of known properties of the t_v and q_v that had to be accommodated by the FSD models. First, these two time series are long-range dependent. This is associated with the important discovery of long-range dependence of packet arrival counts and of packet byte counts in successive equal-length intervals of time, such as 10 ms (Leland, Taqqu, Willinger, and Wilson 1994; Paxson and Floyd 1995). Second, the t_v and q_v are non-Gaussian. The complex non-Gaussian behavior has been demonstrated clearly in important work showing that highly nonlinear multiplicative multi-fractal models can account for the statistical properties of t_v and q_v (Riedi, Crouse, Ribeiro, and Baraniuk 1999; Gao and Rubin 2001). These nonparametric models utilize many coefficients and a complex cascade structure to explain the properties. Third, the statistical properties of the two time series change as c increases (Cao, Cleveland, Lin, and Sun 2002). The arrivals tend toward Poisson and the sizes tend toward independent; there are always long-range dependent components present in the series, but the contributions of the components to the variances of the series goes to zero.

11.4.1   Solving the Non-Gaussian Challenge

The challenge in modeling t_v and q_v is their combined non-Gaussian and long-range dependent properties, a difficult combination that does not, without a simplifying approach, allow parsimonious characterization. We discovered that monotone nonlinear transformations of the inter-arrivals and sizes are very well fitted by parsimonious Gaussian time series, a very simple class of fractional ARIMA models (Hosking 1981) with a small number of parameters. In other words, the transformations and the Gaussian models account for the complex multifractal properties of the t_v and the q_v in a simple way.

11.4.2   The FSD Model Class

Suppose x_v for $v = 1, 2, \ldots$ is a stationary time series with marginal cumulative distribution function $F(x; \phi)$ where ϕ is a vector of unknown parameters. Let $x_v^* = H(x_v; \phi)$ be a transformation of x_v such that the marginal distribution of x_v^* is normal with mean 0 and variance 1. We have $H(x_v; \phi) = G^{-1}(F(x; \phi))$,

where $G(z)$ is the cumulative distribution function of a normal random variable with mean 0 and variance 1. Next we suppose x_v^* is a Gaussian time series and call x_v^* the "Gaussian image" of x_v.

Suppose x_v^* has the following form:

$$x_v^* = \sqrt{1 - \theta}\, s_v + \sqrt{\theta}\, n_v,$$

where s_v and n_v are independent of one another and each has mean 0 and variance 1. n_v is Gaussian white noise, that is, an independent time series. s_v is a Gaussian fractional ARIMA (Hosking 1981)

$$(I - B)^d s_v = \epsilon_v + \epsilon_{u-1}$$

where $Bs_v = s_{u-1}$, $0 < d < 0.5$, and ϵ_v is Gaussian white noise with mean 0 and variance

$$\sigma_\epsilon^2 = \frac{(1 - d)\Gamma^2(1 - d)}{2\Gamma(1 - 2d)}.$$

x_v is a *fractional sum-difference* (*FSD*) time series. Its Gaussian image, x_v^*, has two components. $\sqrt{1 - \theta}s_v$ is the long-range-dependent, or lrd, component; its variance is $1 - \theta$. $\sqrt{\theta}$ is the white-noise component; its variance is θ.

Let $p_{x^*}(f)$ be the power spectrum of the x_v^*. Then

$$p_{x^*}(f) = (1 - \theta)\sigma_\epsilon^2 \frac{4\cos^2(\pi f)}{\left(4\sin^2(\pi f)\right)^d} + \theta$$

for $0 \leq f \leq 0.5$. As $f \to 0.5$, $p_{x^*}(f)$ decreases monotonically to θ. As $f \to 0$, $p_{x^*}(f)$ goes to infinity like $\sin^{-2d}(\pi f) \sim f^{-2d}$, one outcome of long-range dependence. For nonnegative integer lags k, let $r_{x^*}(k)$, $r_s(k)$, and $r_n(k)$ be the autocovariance functions of x_v^*, s_v, and n_v, respectively. Because the three series have variance 1, the autocovariance functions are also the autocorrelation functions. $r_s(k)$ is positive and falls off like k^{2d-1} as k increases, another outcome of long-range dependence. For $k > 0$, $r_n(k) = 0$ and

$$r_{x^*}(k) = (1 - \theta)r_s(k).$$

As $\theta \to 1$, x_v^* goes to white noise: $p_{x^*}(f) \to 1$ and $r_{x^*}(k) \to 0$ for $k > 0$. The change in the autocovariance function and power spectrum are instructive. As θ gets closer to 1, the rise of $p_{x^*}(f)$ near $f = 0$ is always to order f^{-2d}, and the rate of decay of $r_{x^*}(k)$ for large k is always k^{2d-1}; but the ascent of $p_{x^*}(f)$ at the origin begins closer and closer to $f = 0$ and the $r_{x^*}(k)$ get uniformly smaller by the multiplicative factor $1 - \theta$.

11.4.3 Marginal Distributions of q_v and t_v

We model the marginal distribution of t_v by a Weibull with shape λ and scale α, a family with two unknown parameters. Estimates of λ are almost always less than 1. The Weibull provides an excellent approximation of the sample marginal distribution of the t_v except that the smallest 3% to 5% of the sample distribution is truncated to a nearly constant value due to certain network transmission properties.

The marginal distribution of q_v is modeled as follows. While packets less than 40 bytes can occur, it is sufficiently rare that we ignore this and suppose $40 \leq q_v \leq 1500$. First, we provide for A atoms at sizes

$\phi_1^{(s)} \ldots \phi_A^{(s)}$ such as 40 bytes, 512 bytes, 576 bytes, and 1500 bytes that are commonly occurring sizes; the atom probabilities are $\phi_1^{(a)} \ldots \phi_A^{(a)}$. For the remaining sizes, we divided the interval [40, 1500] bytes into C sub-intervals using $C-1$ distinct break points with values that are greater than 40 bytes and less than 1500 bytes, $\phi_1^{(b)}, \ldots \phi_{C-1}^{(b)}$. For each of the C sub-intervals, the size distribution has an equal probability for the remaining sizes (excluding the atoms) in the sub-interval; the total probabilities for the sub-intervals are $\phi_1^{(i)}, \ldots, \phi_C^{(i)}$. Typically, with just 3 atoms at 40 bytes, 576 bytes, and 1500 bytes, and with just 2 break points at 50 bytes and 200 bytes, we get an excellent approximation of the marginal distribution.

11.4.4 Gaussian Images of q_v and t_v

The transformed time series t_v^* and q_v^* appear to be quite close to Gaussian processes. Some small amount of non-Gaussian behavior is still present but it is minor. The autocorrelation structure of these Gaussian images are very well fitted by the FSD autocorrelation structure.

The parameters of the FSD model are the following:

- q_v marginal distribution: A atom probabilities $\phi_i^{(a)}$ at A sizes $\phi_i^{(s)}$; $C-1$ break points $\phi_i^{(b)}$ and C sub-interval probabilities $\phi_i^{(i)}$
- t_v marginal distribution: shape λ and scale α
- q_v^* time dependence: fractional difference coefficient $d^{(q)}$ and white-noise variance $\theta^{(q)}$
- t_v^* time dependence: fractional difference coefficient $d^{(t)}$ and white-noise variance $\theta^{(t)}$.

We found that the $d^{(q)}$ and $d^{(t)}$ do not depend on c; this is based on empirical study, and it is supported by theory. The estimated values are 0.410 and 0.411, respectively. We take the value of each of these two parameters to be 0.41. We found that as c increases, estimates of λ, $\theta^{(q)}$, and $\theta^{(t)}$ all tend toward 1. This means the t_v tend to a independent exponentials, a Poisson process and the q_v tend toward independence. In other words, the statistical models account for the change in the t_v and q_v and c increases that was discussed earlier. We estimated these three parameters and α by partial likelihood methods with $d^{(q)}$ and $d^{(t)}$ fixed to 0.41. The marginal distribution of q_v on a given link does not change with c, but it does change from link to link. To generate traffic, we must specify the atom and sub-interval probabilities. This provides a mean packet size $\mu_{(q)}$, measured in bits/packet.

11.5 Packet Stream Data: Live and Synthetic

We use packet stream data, values of packet arrivals and sizes, in studying the bandwidth estimation problem. They are used as input traffic for queueing simulations. There are two types of streams: live and synthetic. The live streams are from packet traces, data collection from live Internet links. The synthetic streams are generated by the FSD models.

11.5.1 Live Packet Streams

A commonly-used measurement framework for empirical Internet studies results in packet traces (Claffy, Braun, and Polyzos 1995; Paxson 1997; Duffield, Feldmann, Friedmann, Greenberg, Greer, Johnson, Kalmanek, B.Krishnamurthy, Lavelle, Mishra, Ramakrishnan, Rexford, True, , and van der Merwe 2000). The arrival

time of each packet on a link is recorded and the contents of the headers are captured. The vast majority of packets are transported by TCP, so this means most headers have 40 bytes, 20 for TCP and 20 for IP. The live packet traffic is measured by this mechanism over an interval; the time-stamps provide the live inter-arrival times, t_v, and the headers contain information that provides the live sizes, q_v; so for each trace, there is a stream of live arrivals and sizes.

The live stream database used in this presentation consists of 349 streams, 90 sec or 5 min in duration, from 6 Internet links that we name BELL, NZIX, AIX1, AIX2, MFN1, and MFN2. The measured streams have negligible delay on the link input router. The mean number of simultaneous active connections, c, ranges from 49 connections to 18,976 connections. The traffic bit rate, τ, ranges from 1.00 megabits/sec to 348 megabits/sec.

BELL is a 100 megabit/sec link in Murray Hill, NJ, USA connecting a Bell Labs local network of about 3000 hosts to the rest of the Internet. The transmission is half-duplex, so both directions, in and out, are multiplexed and carried on the same link, and a stream is the multiplexing of both directions; but to keep the variable c commensurate for all of the six links, the two directions for each connection is counted as two. In this presentation we use 195 BELL traces, each 5 min in length. NZIX is the 100 megabit/sec New Zealand Internet exchange hosted by the ITS department at the University of Waikato, Hamilton, New Zealand, that served as a peering point among a number of major New Zealand Internet Service Providers at the time of data collection (http://wand.cs.waikato.ac.nz/wand/wits/nzix/2/). All arriving packets from the input/output ports on the switch are mirrored, multiplexed, and sent to a port where they are measured. Because all connections have two directions at the exchange, each connection counts 2 as for BELL. In this presentation we use 84 NZIX traces, each 5 min in length. AIX1 and AIX2 are two separate 622 megabit/sec OC12 packet-over-sonet links, each carrying one direction of the traffic between NASA Ames and the MAE-West Internet exchange. In this presentation we use 23 AIX1 and 23 AIX2 traces, each 90 sec in length. AIX1 and AIX2 streams were collected as part of a project at the National Laboratory for Applied Network Research where the data are collected in blocks of 90 sec (http://pma.nlanr.net/PMA). MFN1 and MFN2 are two separate 2.5 gigabit/sec OC48 packet-over-sonet links on the network of the service provider Metropolitan Fiber Network; each link carries one direction of traffic between San Jose, California, USA and Seattle, Washington, USA. In this presentation we use 12 MFN1 and 12 MFN2 traces, each 5 min in length.

The statistical properties of streams, as we have stated, depend on the connection load c, so it is important that the time interval of a live stream be small enough that c does not vary appreciably over the interval. For any link, there is diurnal variation, a change in c with the time of day due to changes in the number of users. We chose 5 minutes to be the upper bound of the length of each stream to ensure stationarity. BELL, NXIZ, MFN1, and MFN2 streams are 5 min. AIX1 and AIX2 traces are 90 sec because the sampling plan at these sites consisted of noncontiguous 90-sec intervals.

11.5.2 Synthetic Packet Streams

The synthetic streams are arrivals and sizes generated by the FSD models for t_v and q_v. Each of the live streams is fitted by two FSD models, one for the t_v and one for the q_v, and a synthetic stream of 5 min is generated by the models. The generated t_v are independent of the generated q_v, which is what we found in the live data. The result is 349 synthetic streams that match the statistical properties collectively of the live streams.

11.6   Queueing Simulation

We study the bandwidth estimation problem through queueing simulation with an infinite buffer and a first-in-first-out queueing discipline. The inputs to the queues are the arrivals and sizes of the 349 live and 349 synthetic packet streams described in Section 11.5.

For each live or synthetic stream, we carry out 25 runs, each with a number of simulations. For each run we pick a delay δ and a delay probability ω; simulations are carried out to find the QoS bandwidth β, the bandwidth that results in delay probability ω for the delay δ. This also yields a QoS utilization $u = \tau/\beta$. We use 5 delays — 0.001 sec, 0.005 sec, 0.010 sec, 0.050 sec, 0.100 sec — and 5 delay probabilities — 0.001, 0.005, 0.01, 0.02, and 0.05 — employing all 25 combinations of the two delay criteria. For each simulation of a collection, δ is fixed a priori. We measure the queueing delay at the arrival times of the packets, which determines the simulated queueing delay process. From the simulated process we find the delay probability for the chosen δ. We repeat the simulation, changing the trial QoS bandwidth, until the attained delay probability approximately matches the chosen delay probability ω. The optimization is easy because ω decreases as the trial QoS bandwidth increases for fixed δ.

In the optimization, we do not allow the utilization to go above 0.97; in other words, if the true QoS utilization is above 0.97, we set it to 0.97. The reason is that we use the logit scale, $\log(u/(1 - u))$, in the modeling; above about 0.97, the scale becomes very sensitive to model mis-specification and the accuracy of the simulation, even though the utilizations above 0.97 for practical purposes are nearly equal. Similarly, we limit the lower range of the utilizations to be 0.05.

The result of the 25 runs for each of the 349 live and 349 synthetic streams is 25 measurements, one per run, of each of five variables: QoS utilization u, delay δ, delay probability ω, mean number of active connections c, and bit rate τ. The first three variables vary from run to run; the last two variables are the same for the 25 runs for a stream because they measure the stream statistical properties. By design, the range of δ is 0.001 sec to 0.100 sec, and the range of ω is 0.001 to 0.05. The range of the QoS utilizations is 0.05 to 0.97. The two additional variables τ and c, which measure the statistical properties of the streams, are constant across the 25 runs for each stream. c ranges from 49 connections to 18,976 connections, and τ ranges from 1.00 megabits/sec to 348 megabits/sec.

11.7   Model Building: A Bandwidth Formula Plus Random Error

This section describes the process of building the Best-Effort Delay Model, which is the Best-Effort Delay Formula plus random error. The model describes the dependence of the utilization u on the delay δ, the delay probability ω, the traffic bit rate τ, and the expected number of active connections c. The modeling process involves both theory and empirical study, and establishes a basis for the model.

The theoretical basis is queueing theory. The empirical basis is the delay data from the queueing simulations, the measurements of the five variables described in Section 11.6. The following notation is used for the values of these five variables for either the live delay data or the synthetic delay data. δ_j, for $j = 1$ to 5, are the 5 values of the delay in increasing order, and ω_k, for $k = 1$ to 5, are the 5 values of the delay probability in increasing order. u_{ijk} is the QoS utilization for delay δ_j, delay probability ω_k, and stream i, where $i = 1$ to 349. τ_i is the traffic bit rate and c_i is the mean number of active connections, both for stream i.

11.7.1 Strategy: Initial Modeling of Dependence on δ and ω

The structure of the data provides an opportunity for careful initial study of the dependency of the u_{ijk} on δ_j and ω_k. We have 25 measurements of each of these variables for each stream i, and for these measurements, both τ_i and c_i are constant. We start our model building by exploiting this opportunity.

We will consider modeling each stream separately, but hope there will be model consistency across streams that allows simplification. If such simplicity is to occur, it is likely that it will require a monotone transformation of the u_{ijk} because they vary between 0 and 1. So we begin, conceptually, with a model of the form

$$f(u_{ijk}) = g_i(\delta_j, \omega_k) + \epsilon_{ijk}.$$

The ϵ_{ijk} are a sample from a distribution with mean 0. f is a monotone function of u, and g_i is a function of δ and ω; we want to choose f to make g_i as simple as possible, that is, vary as little as possible with i.

11.7.2 Conditional Dependence of u on δ

We start our exploration of the data by taking $f(u) = u$ and suppose that a logical scale for δ is the log. In all cases, we use log base 2 and indicate this by writing $\log_2$ in our formulas. We do not necessarily believe that this identity function for f is the right transformation, but it is helpful to study the data initially on the untransformed utilization scale.

Our first step is to explore the conditional dependence of u_{ijk} on $\log_2(\delta_j)$ given ω_k and the stream i by trellis display (Becker, Cleveland, and Shyu 1996). For each combination of the delay probability ω_k and the stream i, we graph u_{ijk} against $\log_2(\delta_j)$. We did this once for all 349 live streams and once for all 349 synthetic streams. Figure 1 illustrates this by a trellis display for 16 of the live streams. The 16 streams were chosen to nearly cover the range of values of the τ_i. Let $\tau_{(v)}$, for $v = 1$ to 349, be the values ordered from smallest to largest, and take $\tau_{(v)}$ to be the quantile of the empirical distribution of the values of order $v/349$. Then we chose the 16 streams whose ranks v yield orders the closest to the 16 equally-spaced orders from 0.05 to 0.95. On the figure, there are 80 panels divided into 10 columns and 8 rows. On each panel, u_{ijk} is graphed against $\log_2(\delta_j)$ for one value of ω_k and one stream; the strip labels at the top of each panel give the value of ω_k and the rank of the stream. There are 5 points per panel, one for each value of $\log_2(\delta_j)$.

Figure 1 shows a number of overall effects of τ, δ, and ω on u. For each pair of values of ω and τ, there is an increase in u with δ, a strong main effect in the data. In addition, there is an increase with τ for fixed δ and ω, another strong main effect. There is also a main effect for ω, but smaller in magnitude than for the other two variables. The dependence of u on $\log_2(\delta)$ is nonlinear, and changes substantially with the value of τ; as τ increases, the overall slope in u as function of $\log_2(\delta)$ first increases and then decreases. In other words, there is an interaction between $\log_2(\delta)$ and τ. Such an interaction complicates the dependence, so we search further for a transformation f of u that removes the interaction. This pattern occurs when all of the live streams or all of the synthetic streams are plotted in the same way.

There is an interaction between $\log_2(\delta)$ and τ in part because once u is close to 1, where there is little room for change as a function of $\log_2(\delta)$. For this reason, we tried expanding the scale at 1 by taking the function $f(u) = \log_2(1 - u)$. This did not achieve appreciably greater simplicity because nonlinearity and an interaction are still strongly present, but the interaction cause is behavior for smaller values of τ.

The nature of the remaining interaction for $f(u) = \log_2(1 - u)$ suggests that a logit transformation might
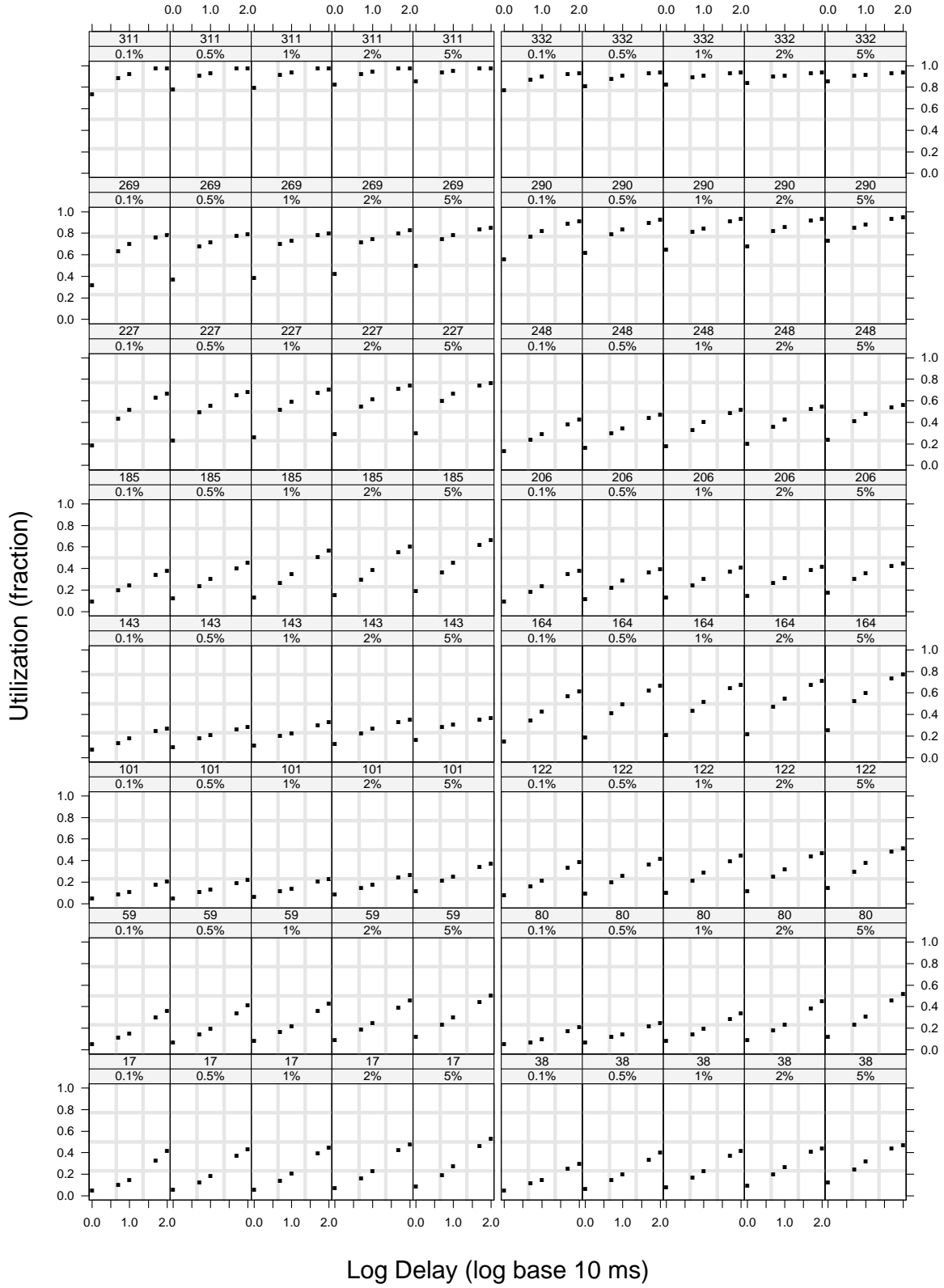
Figure 1: Utilization, u, is graphed against log delay, $\log_2(\delta)$, given the delay probability, ω, and the stream, i.

do better:

$$f(u) = \text{logit}_2(u) = \log_2\left(\frac{u}{1-u}\right).$$

Figure 2 plots $\text{logit}_2(u_{ijk})$ against $\log_2(\delta_j)$ using the same streams and method as Figure 1. The logit function greatly simplifies the dependence. The dependence on $\log_2(\delta)$ is linear. There does not appear to be any remaining interaction among the three variables: $\log_2(\delta)$, τ, and ω. To help show this, 16 lines with different intercepts but the same linear coefficient, have been drawn on the panels. The method of fitting will be described shortly. The lines provide an excellent fit.

11.7.3 Theory: The Classical Erlang Delay Formula

The packet arrivals a_j are not Poisson, although they to tend toward Poisson as c and τ increase. The packet sizes, and therefore the service times, are not independent exponential; the packet sizes have a bounded discrete distribution and are long-range dependent, although they tend to independent as c and τ increase. Still, we will use, as a suggestive case, the results for Poisson arrivals and i.i.d. exponential service times to provide guidance for our model building. Erlang showed that for such a model, the following equation holds (Cooper 1972):

$$\omega = ue^{-(1-u)\beta\delta}.$$

Substituting for $\beta = \tau/u$ and taking the negative log of both sides we have

$$-\log_2(\omega) = -\log_2(u) + \log_2(e)\frac{1-u}{u}\delta\tau. \tag{1}$$

Because ω, which ranges from 0.001 to 0.05, is small in the majority of our simulations compared with u, we have, approximately,

$$-\log_2(\omega) = \log_2(e)\frac{1-u}{u}\delta\tau.$$

Taking logs of both sides and rearranging we have

$$\text{logit}_2(u) = \log_2(\log_2(e)) + \log_2(\tau) + \log_2(\delta) - \log_2(-\log_2(\omega)). \tag{2}$$

So certain aspects of the simplicity of this classical Erlang delay formula occur also in the pattern for our much more statistically complex packet streams. In both cases, $\text{logit}_2(u)$ is additive in functions of τ, δ, and ω, and the dependence is linear in $\log_2(\delta)$.

11.7.4 Conditional Dependence of u on ω

The approximate Erlang delay formula suggests that we try the term $-\log_2(-\log_2(\omega))$, the negative complementary log of ω, in the model. In addition, as we will see in Section 11.9, certain asymptotic results suggest this term as well. We studied the dependence of logit2(u) on $-\log_2(-\log_2(1-\omega))$ for all synthetic and live streams using trellis display in the same way that we studied the dependence on $\log_2(\delta)$. Figure 3 is a trellis plot using the same 16 live streams as in Figure 2. On each panel, $\text{logit}_2(u_{ijk})$ is graphed against $-\log_2(-\log_2(\omega_k))$ for one value of δ_j and one stream.

Figure 3 shows that the guidance from the Erlang formula is on target; logit2(u) is linear in $-\log_2(-\log_2(\omega))$, and the slope remains constant across streams and across different values of δ. To help show this, lines with
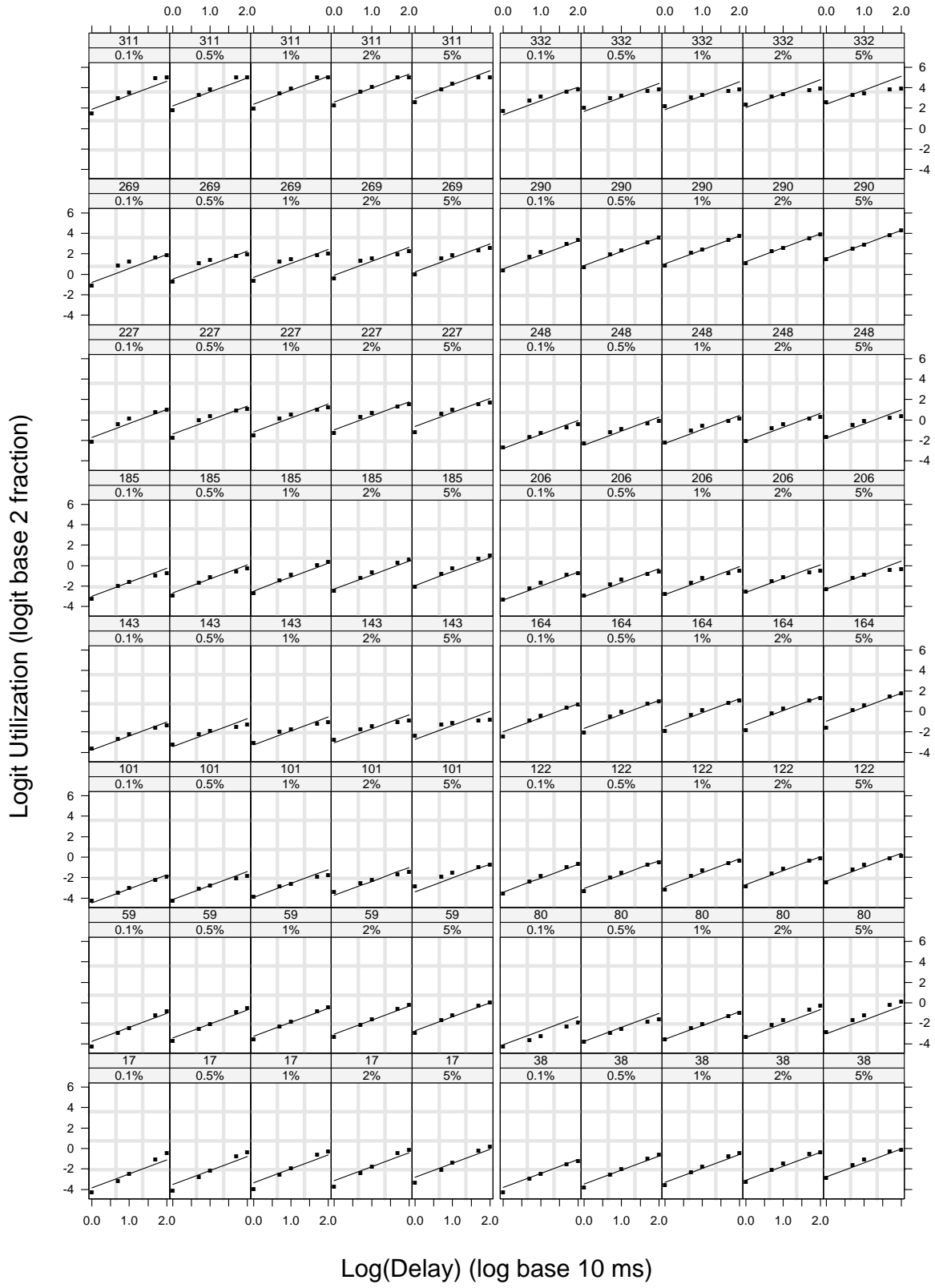
Figure 2: Logit utilization, $\text{logit}_2(u)$, is graphed against log delay, $\log_2(\delta)$, given the delay probability, ω, and the stream, i.
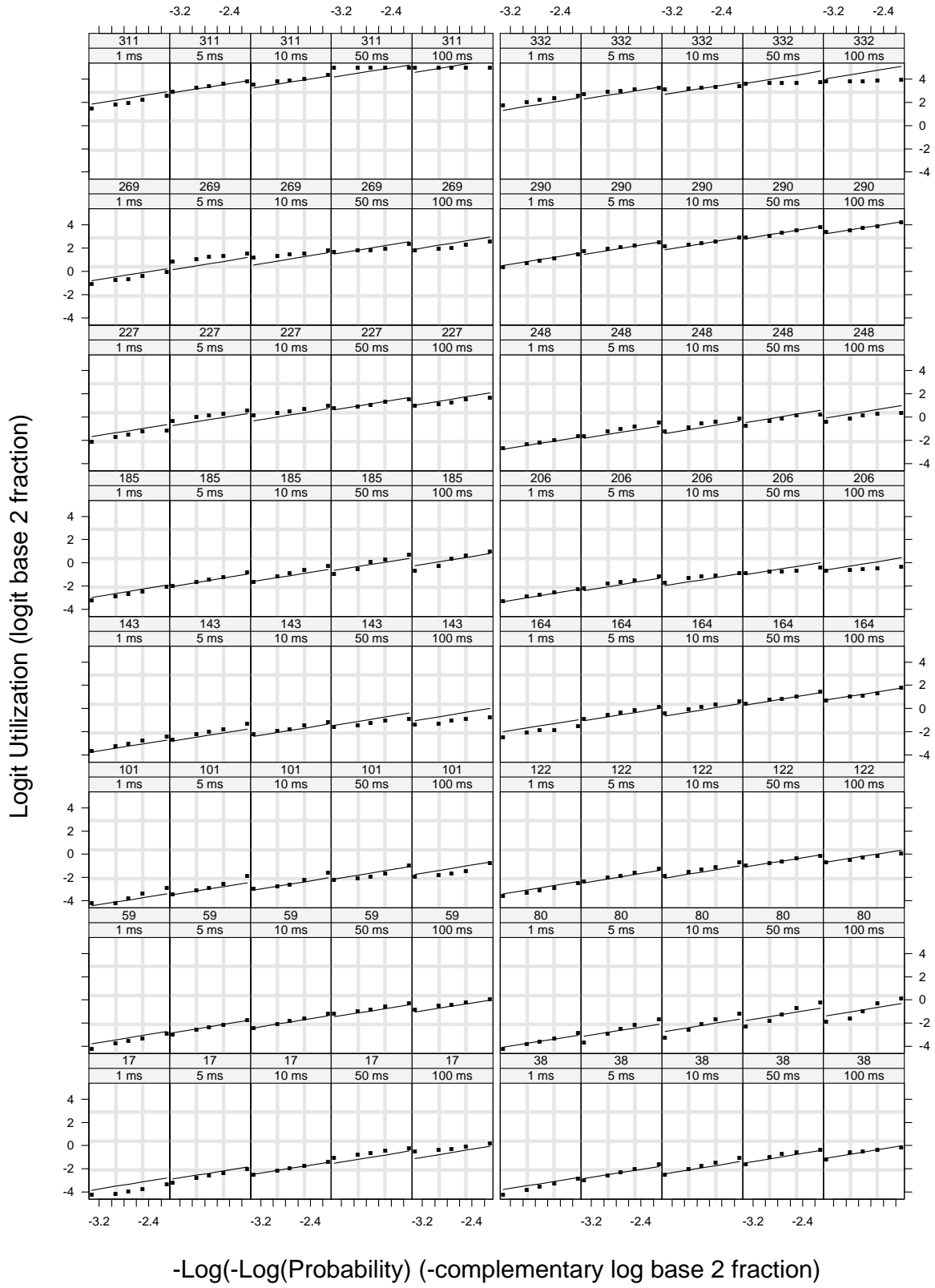
27

Figure 3: Logit utilization, $\text{logit}_2(u)$, is graphed against the negative complementary log of the delay probability, $-\log_2(-\log_2(\omega))$, given the delay, δ, and the stream, i.

the same linear coefficient but different intercepts, have been drawn on the panels. The lines provide an excellent fit except for the errant points for high utilizations observed earlier. The method of fitting will be described shortly. This pattern occurs when all of the live streams or all of the synthetic streams are plotted in the same way.

A Stream-Coefficient Delay Model

The empirical findings in Figures 2 and 3 and the guidance form the Erlang Delay Formula have led to a very simple model that fits the data:

$$, \text{logit}_2(u_{ijk}) = \mu_i + o_\delta \log_2(\delta_j) + o_\omega(-\log_2(-\log_2(\omega_k))) + \epsilon_{ijk} \tag{3}$$

where the ϵ_{ijk} are realizations of an error random variable with mean 0 and median absolute deviation $m(\epsilon)$. The μ_i are stream coefficients; they change with the packet stream i and characterize the statistical properties of the stream.

We fitted the stream-coefficient delay model of Equation 3 twice, once to the 349 live streams and once to the 349 synthetic streams; in other words, we estimated the coefficients μ_i, o_δ, and o_ω twice. Data exploration suggests that the error distribution has longer tails than the normal, so we used the bisquare method of robust estimation (Mosteller and Tukey 1977). The estimates of o_δ and o_ω are

$$\text{Live:} \quad \hat{o}_\delta = 0.411 \ \ \hat{o}_\omega = 0.868$$
$$\text{Synthetic:} \quad \hat{o}_\delta = 0.436 \ \ \hat{o}_\omega = 0.907.$$

The two sets of estimates are very close in the sense that the fitted equation is very close, that is, results in very similar fitted QoS utilizations. For the 16 streams shown in Figures 2 and 3, the lines are drawn using the formula in Equation 3 with the bisquare parameter estimates — $\hat{o}_\delta$, $\hat{o}_\omega$, and $\hat{\mu}_i$.

Because of the long-tailed error distribution, we use the median absolute deviation, m_ϵ, as a measure of the spread. The estimates from the residuals of the two fits are

$$\text{Live:} \quad \hat{m}(\epsilon) = 0.210$$
$$\text{Synthetic:} \quad \hat{m}(\epsilon) = 0.187.$$

The estimates are very small compared with the variation in $\text{logit}_2(u)$. In other words, the stream-coefficient delay model provides a very close fit to the $\text{logit}_2(u_{ijk})$. Of course, this was evident from Figures 2 and 3 because the fitted lines are quite close to the data.

11.7.5   Strategy: Incorporating Dependence on τ and c for Practical Estimation

The coefficient μ_i in the stream-coefficient delay model of Equation 3 varies with the packet stream and reflects how the changing statistical properties of the streams affect the QoS utilization. Part of the simplicity of the model is that a single number characterizes how the statistical properties of a stream affect the QoS bandwidth. However, the model cannot be used as a practical matter for bandwidth estimation because it requires a value of μ, which would not typically be known. If we knew details of the traffic characteristics in detail for the link, for example, if we had FSD parameters, we could generate traffic and run simulations to determine μ and therefore the bandwidth. This might be possible in certain cases, but in general is not feasible.

29

What we must do is start with Equation 3, and find readily-available variables that measure stream statistical properties and can replace μ in the stream-coefficient delay model. We carry out this task in the remainder of this section. Two variables replace μ, the bit rate τ and the mean number of active connections c, with their values of τ_i and c_i for each of our packet streams. We will use both theory and empirical study, as we did for the stream-coefficient delay model, to carry out the model building.

11.7.6   Theory: Fast-Forward Invariance, Rate Gains, and Multiplexing Gains

Figures 2 and 3 show that the QoS utilization u increases with τ. There are two causes: rate gains and multiplexing gains. Because c is positively correlated with τ, u increases with c as well. But τ and c measure different aspects of the load, which is important to the modeling. τ is equal to $c\gamma_b$ where γ_b is the connection bit rate in bits/sec/connection. γ_b measures the end-to-end speed of transfers. c measures the amount of multiplexing. An increase in either increases τ.

First we introduce fast-forwarding. Consider a generalized packet stream with bit rate τ input to a queue without any assumptions about the statistical properties. The packet sizes can be any sequence of positive random variables, and the inter-arrivals can be any point process. Suppose we are operating at the QoS utilization $u = \tau/\beta$ for QoS delay criteria δ and ω. Now for $h > 1$, we speed up the traffic by dividing all inter-arrival times t_v by h. The packet stream has a rate change: the statistical properties of the t_v change only by a multiplicative constant. A rate increase of h increases γ_b by the factor h but not c. The bit rate τ changes to $h\tau$. Suppose we also multiply the bandwidth β by h, so that the utilization u is constant. Then the delay process of the rate-changed packet stream is the delay process for the original packet stream divided by h. That is, if we carried out a simulation with a live or synthetic packet stream, and then repeated the simulation with the rate change, then the delay of each packet in the second simulation would be the delay in the first divided by h. The traffic bit rate, the bandwidth, and the delay process are speeded up by the factor h, but the variation of the packet stream and the queueing otherwise remain the same. If we changed our delay criterion from δ to δ/h, then the QoS utilization u would be the same, which means the QoS bandwidth is $h\beta$. It is as if we videotaped the queueing mechanism in the first simulation and then produced the second by watching the tape on fast forward with the clock on the tape player running faster by the factor h as well; we will call this phenomenon fast-forward invariance.

Let us now reduce some of the speed-up of the fast forwarding. We will divide the t_v by h which increases γ_b by the factor h. But we hold δ fixed and do not decrease by the factor $1/h$. What is the new QoS u that satisfies the delay criteria δ and ω? Since u satisfies the criteria for delay δ/h, we have room for more delay, so u can increase. In other words, a rate increase results in utilization gains for the same δ. This is the rate gain.

Now suppose we hold τ fixed but increase c by the factor $h > 1$. This means that γ_b must be reduced by the factor $1/h$. Now the statistical properties change in other ways due to the increased multiplexing. As we have seen in Section 11.4, the t_v tend toward Poisson and the q_v tend toward independence. The dissipation of the long-range dependence of the packet streams, as well as the tendency of the marginal distribution of t_v toward exponential, tends to decrease the queueing delay distribution and thereby increase the QoS utilization. In other words, there are statistical multiplexing gains.

These theoretical considerations lead us to two important conclusions about modeling. First, we want to be sure that whatever model results, it must obey the principle of fast-forward invariance. Second, it is unlikely that it will be enough to model with just τ. If γ_b were constant across the Internet, τ and c would measure exactly the same thing for our purposes and we would have no need for c beyond τ. But if γ_b changes

substantially, as seems likely, then we will need c as well. Figure 4 shows the 6 medians from the 349 values of $\log_2(\gamma_b)$ for our live streams broken up into 6 groups by the link. The range of the medians is about 4 log base 2 bits/sec/connection, which means that the medians of γ_b change by a factor of 16. One link, NZIX, is appreciably slower than the others.
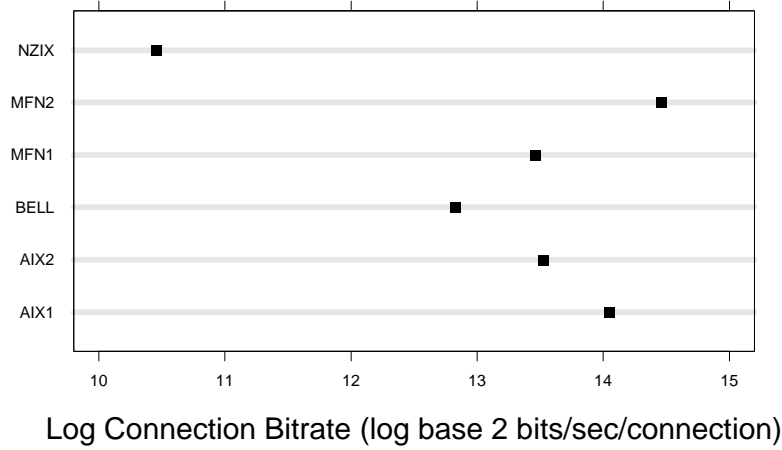

Log Connection Bitrate (log base 2 bits/sec/connection)

Figure 4: Dotplot of median log connection bit rate, $\log_2(\gamma_b)$, for 6 Internet links.

11.7.7 Modeling with τ and c

We will begin by modeling just with τ to see if this can explain the observed utilizations without c. The approximate Erlang delay formula in Equation 2 suggests that the dependence of the stream coefficients on τ is linear in $\log_2(\tau)$. This means the model for $\text{logit}_2(u_{ijk})$ is

$$\text{logit}_2(u_{ijk}) = o + o_\tau \log_2(\tau_i) + o_\delta \log_2(\delta_j) + o_\omega(-\log_2(-\log_2(\omega_k))) + \psi_{ijk}. \tag{4}$$

where the ψ_{ijk} are realizations of an error random variable with mean 0. In our initial explorations for the fit and the residuals we discovered that the spread of the residuals increased with increasing δ_j. So we model the median absolute deviation of the ψ_{ijk} by $m_{\delta_j}(\psi)$, allowing it to change with δ_j.

We fitted the model to the live data and to the synthetic data using bisquare, also accommodating the changing value of $m_{\delta_j}(\psi)$. Figure 5 shows dotplots of the 6 medians of the residuals for the 6 links. There is a clear link effect, mimicking the behavior Figure 4; the two links with the largest and smallest residual medians are the two with the largest and smallest median connection bit rates. The behavior of these extremes is what we would expect. For example, NZIX has the smallest median γ_b, so its bit rate under-predicts the utilization because a stream at NZIX with a certain τ has more multiplexing than average than streams at other links with the same τ, which means the favorable statistical properties push the utilization higher than expected under the model. The same plot for the synthetic streams shows the same effect.

In addition, there is another inadequacy of this first model. Because γ_b is changing, we want the model to obey the principle of fast-forward invariance, but it does not because the estimates of o_τ and o_δ are not equal.

We will enlarge the bandwidth model by adding the variable $\log_2(c)$. Because $\log_2(\tau)$ is used in the initial model, adding $\log_2(c)$ is equivalent to adding γ_b. In doing this we want an equation that obeys fast-forward

31

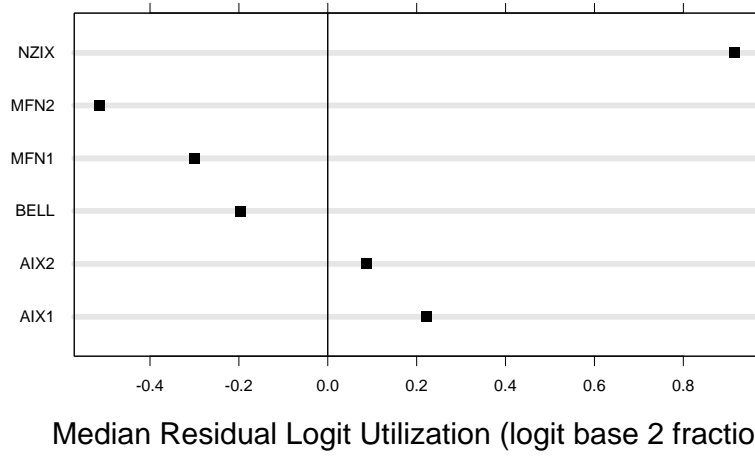Median Residual Logit Utilization (logit base 2 fraction)

Figure 5: Dotplot of link median residuals from the first fit to the logit utilization using only the bit rate τ to characterize stream statistical properties.

invariance: if we hold c fixed, multiply τ by h, and divide δ by h, then we do not want a change in the QoS utilization. This is achieved by the Best-Effort Delay Model:

$$\mathrm{logit}_2(u_{ijk}) = o + o_c \log_2(c_i) + o_{\tau\delta} \log_2(\tau_i \delta_j) + o_\omega(-\log_2(-\log_2(\omega_k))) + \psi_{ijk} \tag{5}$$

where the ψ_{ijk} are error variables with mean 0 and median absolute deviation $m_{\delta_j}(\psi)$. Fast-forward invariance is achieved by entering τ and δ as a product.

We fitted the enlarged of Equation 5 to the live streams and to the synthetic streams using bisquare because our exploration showed that the error distribution has longer tails than the normal; the estimation included a normalization to adjust for the $m_{\delta_j}(\psi)$. The bisquare estimates of o, o_c, $o_{\tau\delta}$, and o_ω are

Live: $\hat{o} = -8.933$ $\hat{o}_c = 0.420$ $\hat{o}_{\tau\delta} = 0.444$ $\hat{o}_\omega = 0.893$

Synthetic: $\hat{o} = -8.227$ $\hat{o}_c = 0.353$ $\hat{o}_{\tau\delta} = 0.457$ $\hat{o}_\omega = 0.952$.

The two sets of estimates are very close in the sense that the fitted equations are close. The estimates of $m_{\delta_j}(\psi)$, the median absolute deviations, $\hat{m}_{\delta_j}(\psi)$, of the residuals, are

Delay :	1 ms	5 ms	10 ms	50 ms	100 ms
Live:	0.211	0.312	0.372	0.406	0.484
Synthetic:	0.169	0.322	0.356	0.380	0.457

Again, the two sets of estimates are close.

It is important to consider whether the added variable c is contributing in a significant way to the variability in $\mathrm{logit}_2(u_{ijk})$, and does not depend fully on the single link NZIX. We used the partial standardized residual plot in Figure 6 to explore this. The standardized residuals of regressing the logit utilization, $\mathrm{logit}_2(u)$, on the predictor variables except $\log_2(c)$, are graphed against the standardized residuals from regressing $\log_2(c)$ on the same variables. The partial regressions are fitted using the final bisquare weights from the full model fit, and the standardization is a division of the residuals by the estimates $\hat{m}_{\delta_j}(\psi)$. Figure 6 shows that $\log_2(c)$ has explanatory power for each link separately and not just across links. We can also see from the plot that there is a remaining small link effect, but a minor one. This is also demonstrated in Figure 7, the same plot as in Figure 5, but for the enlarged model; the horizontal scales on the two plots have been made

32

the same to facilitate comparison. The major link effect is now no longer present in the enlarged model. The result is the same for the same visual display for the synthetic data.

11.7.8 Alternative Forms of the Best-Effort Bandwidth Formula

The Best-Effort Delay Formula of the Best-Effort Delay Model in Equation 5 is

$$\text{logit}_2(u_{ijk}) = o + o_c \log_2(c_i) + o_{\tau\delta} \log_2(\tau_i \delta_j) + o_\omega(-\log_2(-\log_2(\omega_k))). \tag{6}$$

Since $\tau = c\gamma_b$, the formula can be rewritten as

$$\text{logit}_2(u) = o + (o_c + o_{\tau\delta}) \log_2(c) + o_{\tau\delta} \log_2(\gamma_b \delta) + o_\omega(-\log_2(-\log_2(\omega))). \tag{7}$$

In this form we see the action of the amount of multiplexing of connections as measured by c, and the end-to-end connection speed as measured by γ_b. An increase in either results in an increase in the utilization of a link.

11.7.9 Modeling the Error Distribution

As we have discussed, our study of the residuals from the fit of the Best-Effort Delay Model showed that the scale of the residual error distribution increases with the delay. The study also showed that $\log_2(m_\delta)$ is linearly related to $\log_2(\delta)$. From the least squares estimates for the live data, the estimate of the intercept of the regression line is -0.481, the estimate of the linear coefficient of the line is 0.166, and the estimate of the standard error is 0.189. (Results are similar for the synthetic data.)

We also found that when we normalized the residuals by the estimates $\hat{m}_{\delta_j}(\psi)$, the resulting distribution of values is very well approximated by a constant times a t-distribution with 15 degrees of freedom. Because the normalized residuals have a median absolute deviation of 1 and the t_{15} has a median absolute deviation of 0.691, the constant is 0.691^{-1}. We will use this modeling of the error distribution for the bandwidth prediction in Section 11.8.

11.8 Bandwidth Estimation

The Best-Effort Delay Model in Equation 5 can be used to estimate the bandwidth required to meet QoS criteria on delay for best-effort Internet traffic. We describe here a conservative procedure in the sense that the estimated bandwidth will be unlikely to be too small. In doing this we will use the coefficient estimates from the live delay data.

First, we estimate the expected logit utilization, $\ell = \text{logit}_2(u)$ by

$$\hat{\ell} = -8.933 + 0.420 \log_2(c) + 0.444 \underline{\log_2(\tau\delta)} + 0.893(-\log_2(-\log_2(\omega))).$$

On the utilization scale this is

$$\hat{u} = \frac{2^{\hat{\ell}}}{1 + 2^{\hat{\ell}}}.$$

Next we compute a predicted median absolute deviation from the above linear regression:

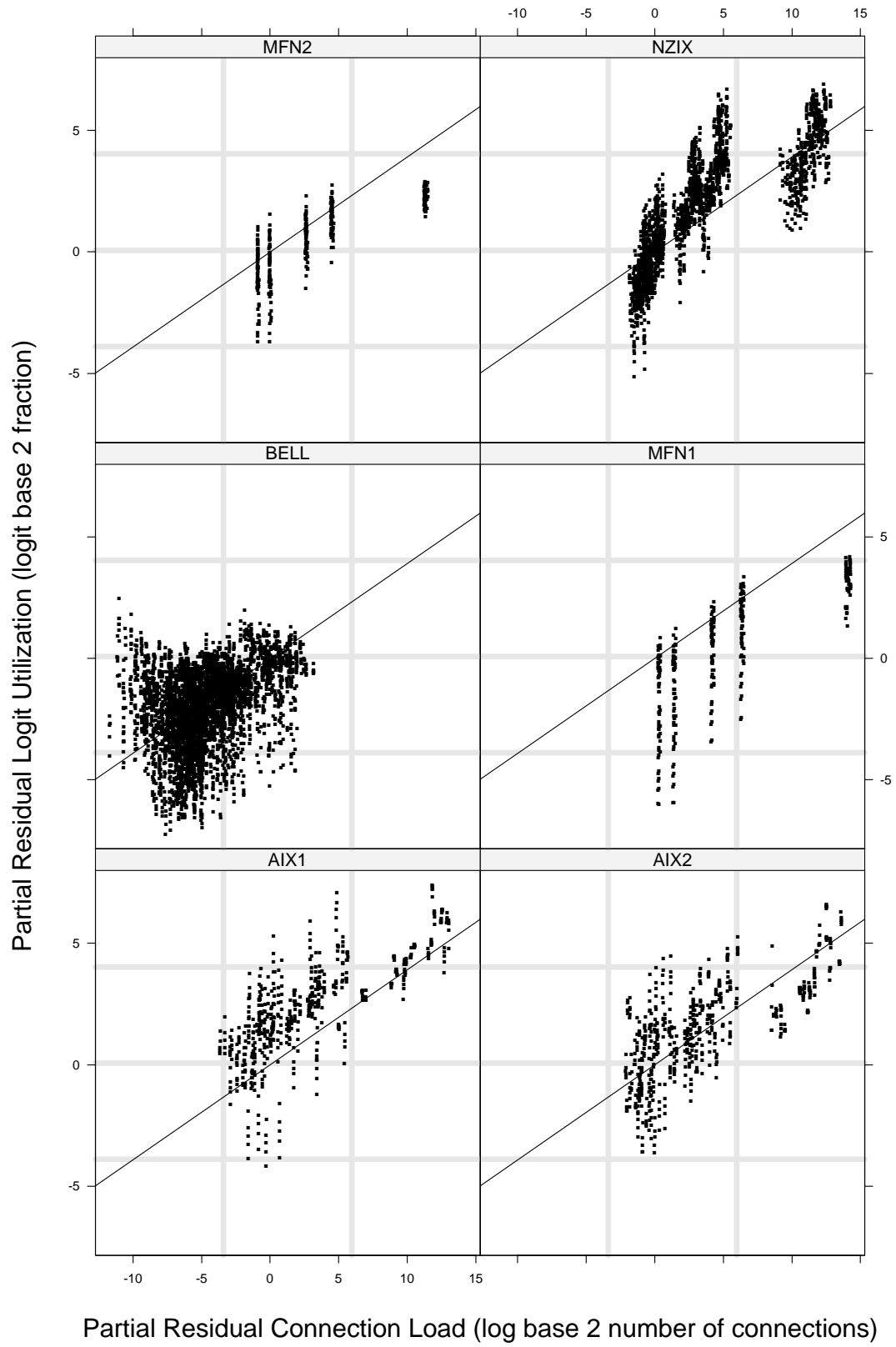$$\hat{m}_\delta(\psi) = 2^{-0.481 + 0.166 \log_2 \delta}.$$

33

Figure 6: A partial residual plot for the explanatory variable $\log_2(c)$ for the Best-Effort Delay Model given each of the 6 Internet links.
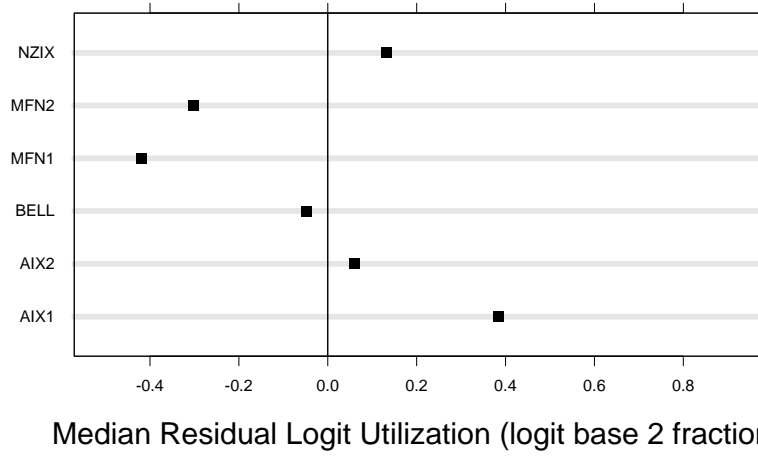
Median Residual Logit Utilization (logit base 2 fraction)

Figure 7: Dotplot of link median residuals for the Best-Effort Delay Model.

Let $t_{15}(p)$ be the quantile of probability p of a t distribution with 15 degrees of freedom. Then the lower limit of a $100(1 - p)\%$ tolerance interval for ℓ is

$$\hat{\ell}(p) = \hat{\ell} - \hat{m}_\delta(\psi) t_{15}(p)/0.691.$$

For $p = 0.05$, $t_{15}(p) = 1.75$, so the lower 95% limit is

$$\hat{\ell}(0.05) = \hat{\ell} - 2.53 \hat{m}_\delta(\psi).$$

The lower 95% limit on the utilization scale is

$$\hat{u}(0.05) = \frac{2^{\hat{\ell}(0.05)}}{1 + 2^{\hat{\ell}(0.05)}}.$$

This process is illustrated in Figure 8. For the figure, γ_b was taken to be 2^{14} bits/sec/connection. On each panel, the values of τ and ω are fixed to the values shown in the strip labels at the tops of the panels, and $\hat{u}$ and $\hat{u}(0.05)$ are both graphed against $\log_2(\delta)$ for δ varying from 0.001 sec to 0.1 sec.

11.9   Other Work on Bandwidth Estimation and Comparison with the Results Here

Bandwidth estimation has received much attention in the literature. The work focuses on queueing because the issue driving estimation is queueing. Some work is fundamentally empirical in nature in that it uses live streams as inputs to queueing simulations or synthetic streams from models that have been built with live streams, although theory can be invoked as well. Other work is fundamentally theoretical in nature in that the goal is to derive mathematically properties of queues, although live data are sometimes used to provide values of parameters so that numerical results can be calculated. Most of this work uses derivations of the delay exceedance probability as a function of an input source to derive the required bandwidth for a given QoS requirement. The delay exceedance probability is equivalent to the our delay probability, where the buffer size is related to the delay by a simple multiplication of the link bandwidth. Since exact calculations of the delay probability are only feasible in special cases, these methods seek an approximate analysis, for example, using asymptotic methods, stochastic bounds, or in some cases, simulations. There has been by far much more theoretical work than empirical.
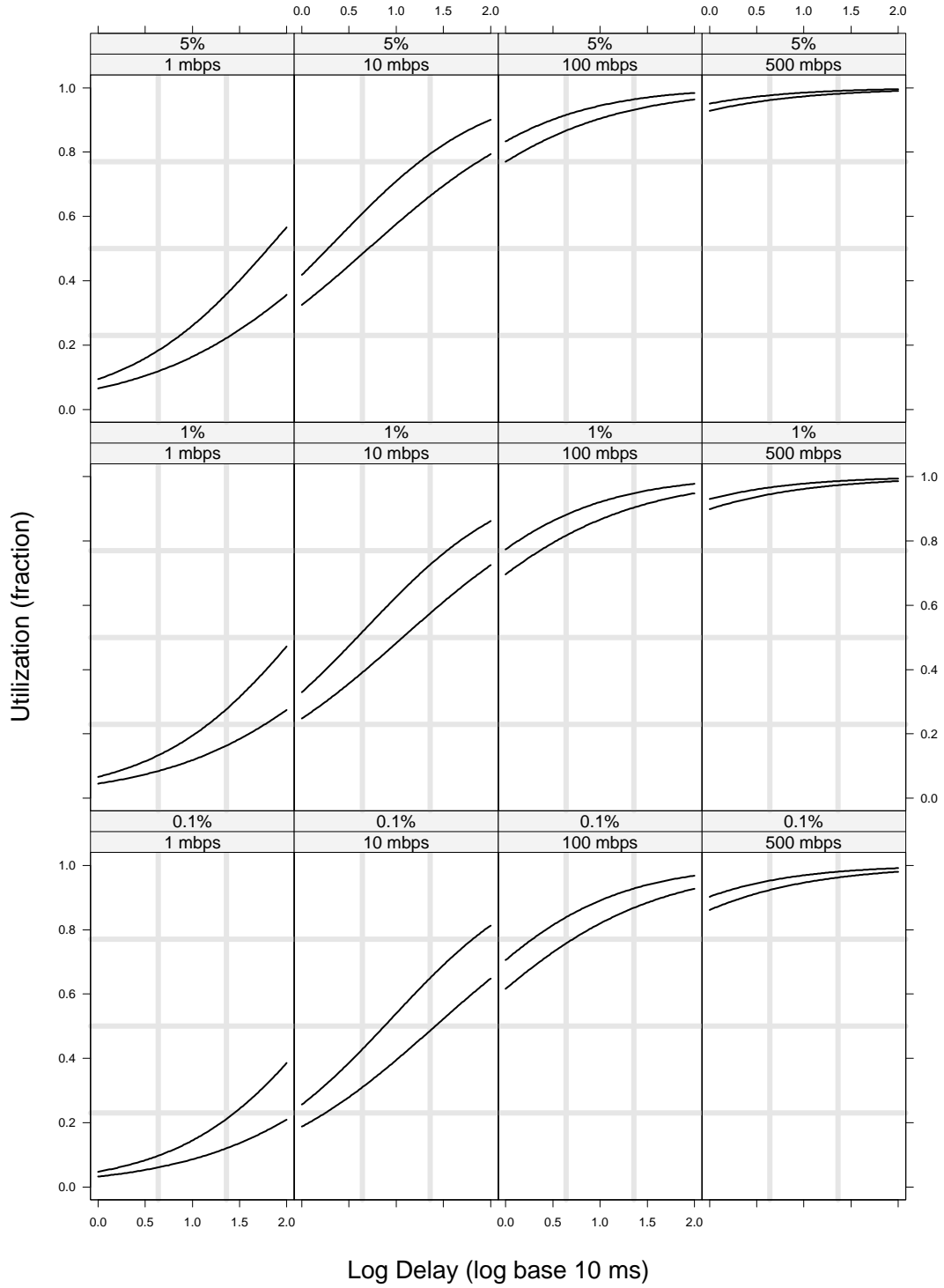
35

Figure 8: The QoS utilization from the Best-Effort Delay Model is graphed against log delay given the bit rate and the delay probability. The bit connection rate γ_b is taken to be 2^{14} bits/sec/connection. The upper curve on each panel estimates the expected values and the lower curve is the minimum values of 95% tolerance intervals.

The attention to the statistical properties of the traffic stream, which have an immense impact on the queueing, receives attention to varying degrees. Those who carry out empirical studies with live streams do so as a guarantee of recreating the properties. Those who carry out studies with synthetic traffic from models must argue for the validity of the models. Much of the theoretical work takes the form of assuming certain stream properties and then, deriving the consequences, so the problem is solved for any traffic that might have these properties. Sometime, though, the problem is minimized by deriving asymptotic results under general conditions.

11.9.1 Empirical Study

Our study here falls in the empirical category but with substantial guidance from theory. To estimate exceedance probabilities, we run simulations of an infinite buffer, FIFO queue with fixed utilization using live packet streams or synthetic streams from the FSD model as the input source.

The tradition for using live Internet streams in a queueing simulation began early in the study of Internet traffic. In a very important study it was shown that long-range dependent traffic results in much greater queue-length distributions (Erramilli, Narayan, and Willinger 1996) than Poisson traffic. This was an important result because it showed that the long-range dependence of the traffic would have a large impact on Internet engineering. In other studies, queueing simulations of both live and altered live traffic are used to study the effect of dependence properties and multiplexing on performance using the average queue length as the performance metric (Erramilli, Narayan, Neidbardt, and Saniee 2000; Cao, Cleveland, Lin, and Sun 2001; Erramilli, Narayan, Neidbardt, and Saniee 2000).

In (Mandjes and Boots 2004), queueing simulations of multiplexed on-off sources are used to study the influence of on and off time distributions on the shape of the loss curve and on performance. To improve the accuracy of the delay probabilities based on simulations, techniques such as importance sampling are also considered (Boots and Mandjes 2001).

One study (Fraleigh, Tobagi, and Diot 2003) first uses live streams, Internet backbone traffic, to validate a traffic model: an extension of Fractional Brownian Motion (FBM), a two-scale FBM process. They do not model the packet process but rather bit rates as a continuous function. Then they derive approximations of the delay exceedance probability from the model, which serves as the basis for their bandwidth estimation. Parameters in the two-scale FBM model that appear in the formula are related to the bitrate τ using the data. This is similar to our process here where we relate the stream coefficients to c and τ. They also used queueing simulations to determine delay exceedance probabilities as a method of validation. We will compare their results and ours at the end of this section.

11.9.2 Mathematical Theory: Effective Bandwidth

A very large number of publications have been written in an area of bandwidth estimation that is referred to as "effective bandwidth". The effective bandwidth of an input source provides a measure of its resource usage for a given QoS requirement, which should lie somewhere between the mean rate and peak rate. Let $A(t)$ be the total workload (e.g., bytes) generated by a source in the interval $[0, t]$, the mathematical definition of the effective bandwidth of the source (Kelly 1996) is

$$\alpha(s, t) = \frac{1}{st} \log \mathrm{E}[e^{sA(t)}] \qquad 0 < s, t < \infty, \tag{8}$$

37

for some space parameter s and time parameter t. For the purpose of bandwidth estimation, the appropriate choice of parameters depends on the traffic characteristics of the source, the QoS requirements, as well as the properties of the traffic with which the source is multiplexed. In the following we shall discuss the effective bandwidth approach to bandwidth estimation based on approximating the delay probability in the asymptotic regime of many sources.

Consider the delay exceedance probability for a First In First Out (FIFO) queue on a link with constant bit rate. In the asymptotic regime of many sources, we are concerned with how the delay probability decays as the size of the system increases. Suppose there are n sources and the traffic generated by the n sources are identical, independent and stationary. The number of sources, n, grows large and at the same time the resources such as the link bandwidth β and buffer sizes scale proportionally so that the delay δ stays constant. Let $\beta = n\beta_0$ for some β_0, and let Q_n be the queueing delay. Under very general conditions, it can be shown that (Botvich and Duffield 1995; Courcoubetis and Weber 1996; Simonian and Guibert 1995; Likhanov and Mazumdar 1998; Mandjes and Kim 2001)

$$\lim_{n \to \infty} -n^{-1} \log \mathrm{P}(Q_n > \delta) = I(\delta, \beta_0), \tag{9}$$

where

$$I(\delta, \beta_0) = \inf_{t>0} \sup_s \left(s\beta_0(\delta + t) - st\alpha(s,t) \right), \tag{10}$$

and is sometimes referred to as the loss curve in the literature. Let (s^*, t^*) be an extremizing pair in Equation (10), then $\alpha(s^*, t^*)$ is the effective bandwidth for the single source as defined in Equation (8), and $n\alpha(s^*, t^*)$ is the effective bandwidth for the n sources. For a QoS requirement of a delay δ and a delay probability w, approximating the delay probability $\mathrm{P}(Q_n > \delta)$ using $\exp(-nI(\delta, \beta_0))$ (Equation 9), the bandwidth required for the n sources can be found by solving the following equation for β

$$s^*\beta(\delta + t^*) - s^*t^*n\alpha(s^*, t^*) = -\log w. \tag{11}$$

This gives

$$\beta = \frac{s^*t^*}{s^*(\delta + t^*)} n\alpha(s^*, t^*) - \frac{\log w}{s^*(\delta + t^*)}, \tag{12}$$

which is the effective bandwidth solution to the bandwidth estimation problem. If the delay $\delta \to \infty$, then the extremizing value of t^* approaches to ∞, and the bandwidth in Equation (12) reduces to

$$n \lim_{t^* \to \infty} \alpha(s^*, t^*),$$

and we recover the classical effective bandwidth definition of a single source $\lim_{t^* \to \infty} \alpha(s^*, t^*)$ for the large buffer asymptotic model (Elwalid and Mitra 1993; Guerin, Ahmadi, and Naghshineh 1991; Kesidis, Walrand, and Chang 1993; Chang and Thomas 1995). If the delay $\delta \to 0$, then the extremizing pair $t^* \to 0$ and $s^*t^* \to \tilde{s}$ for some $\tilde{s}$, the bandwidth in Equation (12) reduces to

$$n \lim_{t^* \to \infty} \alpha(\tilde{s}/t^*, t^*) - \frac{\log w}{\tilde{s}},$$

and we recover the effective bandwidth definition $\lim_{t^* \to \infty} \alpha(\tilde{s}/t^*, t^*)$ for the buffer-less model (Hui 1988).

As we can see, the effective bandwidth solution requires the evaluation of the loss curve $I(\delta, \beta_0)$ (Equation 10). However, explicit form of the loss curve is generally not available. One approach is to derive approximations of the loss curve under buffer asymptotic models, i.e., large buffer asymptotic model ($\delta \to \infty$)

or buffer-less model ($\delta \to 0$), for some classes of input source arrivals. For example, if the source arrival process is Markovian, then for some $\eta > 0$ and ν (Botvich and Duffield 1995)

$$\lim_{\delta \to \infty} I(\delta, \beta_0) - \eta\delta = \nu.$$

If the source arrival is Fractional Brownian Motion with Hurst parameter H, then for some $\nu > 0$ (Duffield 1996)

$$\lim_{\delta \to \infty} I(\delta, \beta_0)/\delta^{2-2H} = \nu.$$

For on-off fluid arrival process, it is shown that as $\delta \to 0$, for some constants $\eta(\beta_0)$ and $\nu(\beta_0)$ (Mandjes and Kim 2001),

$$I(\delta, \beta_0) \sim \eta(\beta_0) + \nu(\beta_0)\sqrt{\delta} + O(\delta),$$

and as $\delta \to \infty$, for some constant $\theta(\beta_0)$ (Mandjes and Boots 2002),

$$I(\delta, \beta_0) \sim \theta(\beta_0)v(\delta),$$

where $v(\delta) = -\log \mathrm{P}(\text{Residual On period} > \delta)$. However, it is found that bandwidth estimation based on buffer asymptotic models suffer practical problems. For the large buffer asymptotic model, the estimated bandwidth could be overly conservative or optimistic due to the fact that it does not take into account the statistical multiplexing gain (Choudhury, Lucantoni, and Whitt 1994; Knightly and Shroff 1999). For the buffer-less model, there is a significant utilization penalty in the estimated bandwidth (Knightly and Shroff 1999) since results indicate that there is a significant gain even with a small buffer (Mandjes and Kim 2001).

Another approach proposed by (Courcoubetis, Siris, and Stamoulis 1999; Courcoubetis and Siris 2001) is to numerically evaluate the loss curve $I(\delta, \beta_0)$. First, they evaluate the effective bandwidth function $\alpha(s,t)$ (Equation 8) empirically based on measurements of traffic byte counts in fixed size intervals, and then they obtain the loss curve (Equation 10) using numeric optimizing procedures with respect to the space parameter s and the time parameter t. As examples, they applied this approach to estimate bandwidth where the input source is a Bellcore Ethernet WAN streams or streams of incoming IP traffic over the University of Crete's wide area link. Their empirical approach is model free in the sense that it does not require a traffic model for the input source and all evaluations are based on traffic measurements. However, their approach is computational intensive, not only because the effective bandwidth function $\alpha(s,t)$ has to be evaluated for all time parameters t, but also because the minimization with respect to t is non-convex (unlike the maximization in the space parameter s) and thus difficult to perform numerically (Gibbens and Teh 1999; Kontovasilis, Wittevrongel, Bruneel, Houdt, and Blondia 2002).

In the effective bandwidth approach, one typically approximates the buffer exceedance probability based on its logarithmic asymptote. For example, in the asymptotic regime of many sources, using Equation (9), one can approximate

$$\mathrm{P}(Q_n > \delta) \approx \exp(-nI(\delta, \beta_0)). \tag{13}$$

An improved approximation can be found by incorporating a pre-factor, i.e.

$$\mathrm{P}(Q_n > \delta) \approx K(n, \delta, \beta_0)\exp(-nI(\delta, \beta_0)).$$

Using the Bahadur-Rao theorem, such approximation has been obtained for the delay exceedance probability in the infinite buffer case as well as the cell loss ratio in the finite buffer case that has the same logarithmic asymptote but a different pre-factor (Likhanov and Mazumdar 1998).

11.9.3 Theory: Other Service Disciplines

Some authors have investigated service disciplines other than FIFO such as general processor sharing (li Zhang, Towsley, and Kurose 1994) and priority queue (Berger and Whitt 1998). Although TCP is the most dominant protocol in today's Internet, we do not consider the effect of the TCP feedback control mechanism since the link we sought for estimating a bandwidth is not a bottleneck link. To account for the TCP feedback control, other authors have studied characteristics of bandwidth sharing for elastic traffic and investigated the bandwidth estimation problem for such traffic (de Veciana, Lee, and Konstantopoulos 2001; Fredj, Bonald, Proutiere, Regnie, and Roberts 2001). Again other authors have considered regulated input traffic such as those from a leaky bucket (Elwalid, Mitra, and Wentworth 1995; Presti, Zhang, Kurose, and Towsley 1999; Kesidis and Konstantopoulos 2000; Cheng-Shang Chang and Song 2001).

11.9.4 Theory: Direct Approximations of the Delay Probability

Besides approximating the delay exceedance probability using the effective bandwidth approach, some authors have considered direct approximations for some special classes of input traffic models. For example, for a Markov modulated fluid source, the delay probability can be more accurately expressed as a single exponential with a prefactor K determined from the loss probability in a buffer-less multiplexer as estimated by Chernoff's theorem (Elwalid, Heyman, Lakshman, Mitra, and Weiss 1995). For an aggregate Markov modulated fluid source, the delay probabilities can be approximated by a sum of exponentials (Shroff and Schwartz 1998). For a Gaussian process, a tight lower bound of the delay probability can be obtained using maximum-variance based approaches (Norros 1994; Knightly 1997; Kim and Shroff 1998; Fraleigh, Tobagi, and Diot 2003). These expressions can be used in place of Equation (13) to derive the required bandwidth for a QoS requirement. Readers are referred to (Knightly and Shroff 1999) for a nice overview and comparison of these approaches as well as the aforementioned effective bandwidth approach for bandwidth estimation.

11.9.5 Theory: Queueing Distributions

We now discuss implications of our stream coefficient delay formula and Best-Effort Delay Formula and its relation to some previous work. The stream coefficient delay model in Equation 3 implies that for each stream i,

$$w \approx \mathrm{P}(Q_i > \delta) \approx \exp\left(-\log 2 \cdot 2^{\frac{\mu_i}{o_w}}\left(\frac{u}{1-u}\right)^{-\frac{1}{o_w}}\delta^{\frac{o_\delta}{o_w}}\right)$$

for stream coefficient μ_i and regression coefficients o_w, o_δ. This suggests the tail distribution of queueing delay is Weibull with shape parameter $o_\delta o_w^{-1}$. The Weibull form is consistent with the FBM traffic model (and also the two-scale FBM model), but there the shape parameter is $2 - 2H$. Notice that $o_\delta o_w^{-1}$ from our analysis is 0.52 for the real data and 0.42 for the synthetic data which is quite different from the shape parameter computed from $2 - 2H = 0.18$. If the bitrate per connection γ_b is a fixed constant, the Best-Effort Delay Formula in Equation 5 implies that for some constant o'

$$w \approx \mathrm{P}(Q_i > \delta) \approx \exp\left(-\log 2 \cdot 2^{\frac{o'}{o_w}}\tau_i^{\frac{o_c + o_{\tau\delta}}{o_w}}\left(\frac{u}{1-u}\right)^{-\frac{1}{o_w}}\delta^{\frac{o_\delta}{o_w}}\right).$$

If $o_c + o_{\tau\delta} = o_w$ and the traffic bitrate τ_i is a multiple of τ, i.e. $\tau_i = n_i\tau$, then the above approximation is consistent with the effective bandwidth result with many sources asymptotics (Equation 9). In our empirical

analysis, we have found $o_{\tau\delta} + o_c$ and o_w to be quite close with the ratio $(o_{\tau\delta} + o_c)o_w^{-1}$ being 0.97 for real data and 0.85 for synthetic data. One of the reasons that this ratio is not 1 is possibly due to the fact that (Equation 9) is an asymptotic formula.

11.9.6 Comparison of The Results Presented Here with Other Work

The work presented in this article has resulted in a simple formula for bandwidth estimation. At the same time, validation has been extensive, permeating all areas of the work. Validation is carried out in two ways: empirically and theoretically.

The large number of papers in the area of effective bandwidth and other theoretical work cited above have yielded much insight. The work has posited traffic stream models and investigated the resulting mathematical properties. However, for best-effort Internet traffic, there has not been extensive study to determine whether some posited model accurately describes the stream statistical properties. Nor has there been extensive empirical work in the form of extensive empirical queueing simulations to determine whether queueing results for best-effort traffic fit the theory. Consequently, the simple Best-Effort Delay Formula, not readily derivable without a hint of the final results, was not uncovered.

The interesting paper cited above using the two-scale FBM model surely took great pains to validate the model (Fraleigh, Tobagi, and Diot 2003). One problem with this approach, modeling traffic bit flow as a fluid rather than the packet process as it appears on the link, is that the Gaussian assumption does not take hold until the level of aggregation is quite high. Consequently, the FBM model is not a good approximation until the traffic rate is 50 megabits/sec and above. So their bandwidth estimation model is not a validated one below 50 megabits/sec. By contrast, our Best-Effort Delay Model is valid to as low as 1 megabit/sec. But the ensuing methods used in (Fraleigh, Tobagi, and Diot 2003) to find the QoS bandwidth require a series of approximations and a worst-case empirical method in the estimation of parameters. There appears to have been little checking of these approximations. The bandwidth results appear to us to be inaccurate, possibly arising from some of the approximations. First, as the bit rate increases up to 1 gigabit/sec, the utilization appears to stabilize at values less than 1, substantially so in some cases. As our theoretical discusses of rate gains and multiplexing gains demonstrates, the utilization must increases to 1 as the bitrate increases. This is the case for our Best-Effort Delay Model. For example, the utilization for a delay of 10 ms, a probability of 0.01, and a bit rate of 1 gigabit/sec, is 90% from Figure 8 of (Fraleigh, Tobagi, and Diot 2003) but is 98% for our model. In addition, the model in (Fraleigh, Tobagi, and Diot 2003) works simply with the bit rate rather than decomposing into the number of active connections times the bit rate per connection and using two variables, as is done in the Best-Effort Delay Model here. As we have demonstrated, theoretically and empirically, the bit rate is not sufficient to account for the utilization since a fast network and one with a high traffic connection load must be distinguished.

11.10 Results and Discussion

11.10.1 Problem Formulation

Suppose the packet stream — packet arrival times and sizes — arriving for transmission on an Internet link is best-effort traffic with bit rate τ bits/sec and number of simultaneous active connections, c. Suppose the link input buffer is large enough that packet loss is negligible. Our goal is to estimate the quality-of-service (QoS) bandwidth β in bits/sec, or equivalently, the QoS utilization $u = \tau/\beta$, that satisfies QoS criteria for

the packet queueing delay in the link input buffer. The criteria are a delay, δ sec, and the probability, ω, that the delay for a packet exceeds δ.

11.10.2   Other Work on the Problem

There is a wide literature on the bandwidth estimation problem. Much of it is theoretical — mathematical results that derive properties of queueing systems. A smaller literature is empirical in nature, based on simulations with packet stream inputs from measurements on live links or from models for traffic. The classical Erlang Delay Formula provides a simple formula that can be used for estimation of traffic streams that in theory have Poisson arrivals and i.i.d. exponential sizes. Best-effort traffic is much more complex, it is nonlinear and long-range dependent, and there has been to date no simple, validated formula for it.

11.10.3   Principal Result: The Best-Effort Delay Model

The principal result of this paper is a statistical model that provides a simple, validated formula for the estimation of bandwidth for best-effort traffic in the same way that the Erlang Delay Formula does for the Poisson-exponential case. The model has been validated through extensive empirical study and through consistency with certain theoretical properties of queueing.

The model consists of the Best-Effort Delay Formula plus random variation:

$$\text{logit}_2(u) = o + o_c \log_2(c) + o_{\tau\delta} \log_2(\tau\delta) + o_\omega(-\log_2(-\log_2(\omega))) + \psi.$$

ψ is a random error variable with mean 0 and median absolute deviation $m_\delta(\psi)$, which depends on δ; $\log_2$ is the log base 2; and $\text{logit}_2(u) = \log_2(u/(1-u))$. The distribution of $0.691\psi/m_\delta(\psi)$ is a t-distribution with 15 degrees of freedom. Estimates of the coefficients of the model are

$$\hat{o} = -8.933 \ \ \hat{o}_c = 0.420 \ \ \hat{o}_{\tau\delta} = 0.444 \ \ \hat{o}_\omega = 0.893.$$

$m_\delta(\psi)$ is modeled as a function of δ: $\log_2(m_\delta(\psi))$ is a linear function of $\log_2(\delta)$ plus random variation. The estimate of the intercept of the line is –0.481, the estimate of the linear coefficient of the line is 0.166, and the estimate of the standard error is 0.189. The bit rate τ is equal to $c\gamma_b$ where γ_b is the connection bit rate in bits/sec/connection. So the Best-Effort Delay Formula can also be written as

$$\text{logit}_2(u) = o + (o_c + o_{\tau\delta}) \log_2(c) + o_{\tau\delta} \log_2(\gamma_b\delta) + o_\omega(-\log_2(-\log_2(\omega))).$$

In this form we see the action of the amount of multiplexing of connections as measured by c, and the end-to-end connection speed as measured by γ_b. An increase in either results in an increase in the utilization of a link.

The Best-Effort Delay Model is used to estimate the bandwidth required to carry best-effort traffic given δ, ω, τ, and c. The QoS logit utilization is estimated by

$$\hat{\ell} = -8.933 + 0.420 \log_2(c) + .444 \log_2(\tau\delta) + 0.893(-\log_2(-\log_2(\omega))),$$

so the QoS utilization is estimated by

$$\hat{u} = \frac{2^{\hat{\ell}}}{1 + 2^{\hat{\ell}}}.$$

The corresponding estimated bandwidth is $\tau/\hat{u}$. For such an estimate there is a 50% chance of being too large and a 50% chance of being too small. We might however, use a more conservative estimate that provides a much smaller chance of too little bandwidth. Let

$$\hat{m}_\delta(\psi) = 2^{-0.481+0.166\log_2(\delta)}$$

be the estimate of $m(\delta)$. Let $t_{15}(p)$ be the lower $100p\%$ percentage point of a t-distribution with 15 degrees of freedom, where p is small, say 0.05. Let

$$\hat{\ell}(p) = \hat{\ell} - \hat{m}(\delta)t_{15}(p)/0.691$$

Then

$$\hat{u}(p) = \frac{2^{\hat{\ell}(p)}}{1 + 2^{\hat{\ell}(p)}},$$

is a conservative utilization estimate, the lower limit of a $100p\%$ tolerance interval for the QoS utilization. The corresponding estimated bandwidth is $\tau/\hat{u}(p)$.

11.10.4 Methods

The Best-Effort Delay Model was built, in part, from queueing theory. Certain predictor variables were suggested by the Erlang Delay Formula. Theory prescribes certain behavior as τ, c, or γ_b increase, resulting in rate gains, multiplexing gains, or fast-forward invariance, and the model was constructed to reproduce the behavior.

The Best-Effort Delay Model was built, in part, from results of queueing simulations with traffic stream inputs of two types, live and synthetic. The live streams are measurements of packet arrivals and sizes for 349 intervals, 90 sec or 5 min in duration, from 6 Internet links. The synthetic streams are arrivals and sizes generated by recently developed FSD time series models for the arrivals and sizes of best-effort traffic. Each of the live streams was fitted by two FSD models, one for the inter-arrivals and one for the sizes, and a synthetic stream of 5 min was generated by the models. The generated inter-arrivals are independent of the generated sizes, which is what we found in the live data. The result is 349 synthetic streams that match the statistical properties collectively of the live streams. For each live or synthetic stream, we carried out 25 runs, each with a number of simulations. For each run we picked a delay δ and a delay probability ω; simulations were carried out to find the QoS bandwidth β, the bandwidth that results in delay probability ω for δ. This also yields a QoS utilization $u = \tau/\beta$. We used 5 delays — 0.001 sec, 0.005 sec, 0.010 sec, 0.050 sec, 0.100 sec — and 5 delay probabilities — 0.001, 0.005, 0.01, 0.02, and 0.05 — employing all 25 combinations of the two delay criteria. The queueing simulation results in delay data: values of five variables — QoS utilization u, delay δ, delay probability ω, the mean number of active connections of the traffic c, and the traffic bit rate τ. The delay data were used in the model building.

11.10.5 Validity and Applicability

Extensive data exploration with visualization tools, some shown here, demonstrate that the Best-Effort Delay Model fits the simulation delay data. This, of course, is necessary for the model to be valid. In addition, validity is supported by the model reproducing the theoretical queueing properties as just discussed.

The validity of the Best-Effort Delay Model depends on the validity of the traffic streams used as inputs to the queueing simulation; that is, the packet streams must reproduce the statistical properties of best-effort streams. Of course, the live streams of the study do so because they are best-effort traffic. Extensive validation has shown that the FSD models, used to generate the packet streams here, provide excellent fits to best-effort packet streams when c is above about 64 connections, which for a link where γ_b is about 2^{14} bits/sec/connection means τ is above about 1 megabit/sec. For this reason, only traffic streams with τ greater than this rate are used in the study, and the Best-Effort Delay Model is valid above this rate.

The results are only valid for links with a large enough buffer that the packet loss is negligible. We have used open-loop study, which does not provide for the TCP feedback that occurs when loss is significant. This restriction also holds for the other work on bandwidth estimation cited here.

There is also a practical restriction on applicability. We have taken the range of our study to include traffic bit rates as low as about 1 megabit/sec. We have done this simply because we can do so and achieve valid results. But even for the least stringent of our delay criteria, $\delta = 0.1$ sec delay and $\omega = 0.05$ delay probability, the utilizations are low for rates in the range of 1 megabit/sec to 5 megabits/sec. This might well be judged to be too small a utilization to be practical; if so, it might mean that the what must be sacrificed is the negligible packet loss, which means that a QoS study at the very low traffic bit rates needs to take account of TCP feedback.

One outcome of the dependence of the bandwidth estimation on the traffic statistics is that our solution for best-effort traffic would not apply to other forms of Internet traffic that do not share the best-effort statistical properties. One example is voice traffic.

Finally, the Best-Effort Delay Model provides an estimation of bandwidth in isolation without considering other network factors. A major factor in network design is link failures. Redundancy needs to be built into the system. So an estimate of bandwidth from the model for a link based on the normal link traffic may be reduced to provide the redundancy. But the model still plays a role because the bandwidth must be chosen based on link traffic, but now the traffic in the event of a failure elsewhere.

Bibliography

Becker, R. A., W. S. Cleveland, and M. J. Shyu (1996). The Design and Control of Trellis Display. *Journal of Computational and Statistical Graphics 5*, 123–155.

Berger, A. W. and W. Whitt (1998, August). Effective bandwidths with priorities. *IEEE/ACM Transactions on Networking 6*(4), 447–460.

Boots, N. and M. Mandjes (2001). Fast simulation of a queue fed by a superposition of many heavy-tailed sources. *Probability in the Engineering and Informational Sciences 16*, 205–232.

Botvich, D. D. and N. G. Duffield (1995). Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems 20*, 293–320.

Cao, J., W. S. Cleveland, D. Lin, and D. X. Sun (2001). On the Nonstationarity of Internet Traffic. *ACM SIGMETRICS 29*(1), 102–112.

Cao, J., W. S. Cleveland, D. Lin, and D. X. Sun (2002). Internet Traffic Tends *Toward* Poisson and Independent as the Load Increases. In C. Holmes, D. Denison, M. Hansen, B. Yu, and B. Mallick (Eds.), *Nonlinear Estimation and Classification*, pp. 83–109. New York: Springer.

Cao, J., W. S. Cleveland, and D. X. Sun (2004). Fractional Sum-Difference Models for Open-Loop Generation of Internet Packet Traffic. Technical report, Bell Labs, stat.bell-labs.com.

Cao, J. and K. Ramanan (2002). A Poisson Limit for Buffer Overflow Probabilities. In *IEEE INFOCOMM*, Volume 2, pp. 994 – 1003.

Chang, C.-S. and J. Thomas (1995). Effective bandwidth in high speed digital networks. *IEEE Journal on Selected Areas in Communications 13*, 1091–1100.

Cheng-Shang Chang, Y.-m. C. and W. T. Song (2001). On the performance of multiplexing independent regulated inputs. In *Proc. ACM Sigmetrics 2001/Performance 2001*, pp. 184–193.

Choudhury, G. L., D. M. Lucantoni, and W. Whitt (1994). On the effectiveness of effective bandwidths for admission control in atm networks. In *Proceedings of the 14th International Teletraffic Congress (ITC-14)*, North Holland, pp. 411–420. Elsevier Science B. V.

Claffy, K., H.-W. Braun, and G. Polyzos (1995). A Parameterizable Methodology for Internet Traffic Flow Profiling. *IEEE Journal on Selected Areas in Communications 13*, 1481–1494.

Cooper, R. B. (1972). *Queueing Theory*. New York: Macmillan.

Courcoubetis, C. and V. A. Siris (2001). Procedures and tools for analysis of network traffic measurements. Technical report.

Courcoubetis, C., V. A. Siris, and G. D. Stamoulis (1999). Application of the many sources asymptotic and effective bandwidths to traffic engineering. *Telecommunication Systems 12*, 167–191.

Courcoubetis, C. and R. Weber (1996). Buffer overflow asymptotics for a buffer handling many traffic sources. *Journal of Applied Probability 33*, 886–903.

de Veciana, G., T.-J. Lee, and T. Konstantopoulos (2001, February). Stability and performance analysis of networks supporting elastic services. *IEEE/ACM Transactions on Networking 9*(1), 2–14.

Duffield, N., A. Feldmann, J. Friedmann, A. Greenberg, R. Greer, T. Johnson, C. Kalmanek, B.Krishnamurthy, D. Lavelle, P. Mishra, K. Ramakrishnan, J. Rexford, F. True, , and J. van der Merwe (2000). Measurement and Analysis of IP Network Usage and Behavior. *IEEE Communications Magazine May*, 82–86.

Duffield, N. G. (1996). Economies of scale in queues with sources having power-law large deviations scalings. *Journal of Applied Probability 33*, 840–857.

Elwalid, A., D. Heyman, T. V. Lakshman, D. Mitra, and A. Weiss (1995, Aug). Fundamental bounds and approximations for atm multiplexers with applications to video teleconferencing. *IEEE Journal of Selected Areas in Communications 13*(6), 1004–1016.

Elwalid, A. and D. Mitra (1993). Effective bandwidth of general markovian traffic sources and admission control of high speed networks. *IEEE/ACM Transactions on Networking 1*, 329–343.

Elwalid, A., D. Mitra, and R. H. Wentworth (1995). A new approach for allocating buffers and bandwidth to heterogeneous regulated traffic in an atm node. *IEEE Journal of Selected Areas in Communications 13*(6), 1115–1127.

Erramilli, A., O. Narayan, A. Neidbardt, and I. Saniee (2000). Performance Impacts of Multi-Scaling in Wide Area TCP/IP traffic. In *Proceedings of IEEE Infocom*.

Erramilli, A., O. Narayan, and W. Willinger (1996). Experimental Queueing Analysis with Long-Range Dependent Packet Traffic. *IEEE/ACM Transactions on Networking 4*, 209–223.

Fraleigh, C., F. Tobagi, and C. Diot (2003). Provisioning ip backbone networks to support latency sensitive traffic. In *IEEE Infocom*.

Fredj, S. B., T. Bonald, A. Proutiere, G. Regnie, and J. W. Roberts (2001, August). Statistical bandwidth sharing: A study of congestion at flow level. In *SigComm'01*, San Diego, California, USA, pp. 111–122.

Gao, J. and I. Rubin (2001). Multiplicative Multifractal Modeling of Long-Range-Dependent Network Traffic. *International Journal of Communications Systems 14*, 783–201.

Gibbens, R. J. and Y. C. Teh (1999, June). Critical time and space scales for statistical multiplexing. In *Proceedings of the 16th Internaltional Teletraffic Congress (TC 16)*, Edinburgh, UK, pp. 87–96.

Guerin, R., H. Ahmadi, and M. Naghshineh (1991, Sep). Equivalent capacity and its applications to bandwidth allocation in high-speed networks. *IEEE Journal on Selected Areas in Communications 9*(7), 968–981.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika 68*, 165–176.

http://pma.nlanr.net/PMA. Passive Measurement and Analysis Project at NLANR.

http://wand.cs.waikato.ac.nz/wand/wits/nzix/2/. NZIX-II Trace Data.

Hui, J. Y. (1988). Resource allocation for broadband networks. *IEEE Journal Selected Areas in Communications 6*, 1598–1608.

Iyer, S., S. Bhattacharyya, N. Taft, N. McKeown, and C. Diot (2003). An approach to alleviate link overload as observed on an IP backbone. In *IEEE Infocom*, San Francisco.

Kelly, F. (1996). Notes on effective bandwidths. In S. Z. F.P. Kelly and I. Ziedins (Eds.), *Stochastic Networks: Theory and Applications*, Number 4 in Royal Statistical Society Lecture Notes Series, pp. 141–168. Oxford University Press.

Kesidis, G. and T. Konstantopoulos (2000, Jan). Worst-case performance of a buffer with independent shaped arrival processes. *IEEE Communications Letters* , 26–28.

Kesidis, G., J. Walrand, and C.-S. Chang (1993, August). Effective bandwidths for multiclass markov fluids and other atm sources. *IEEE/ACM Transactions on Networking 1*(3), 424–428.

Kim, H. and N. Shroff (1998, Oct). A central limit theorem based approach to analyze queue behavior in atm networks. *IEEE/ACM Transactions on Networking 6*, 659–671.

Knightly, E. (1997, June). Second moment resouce allocation in multi-service networks. In *Proc. ACM Sigmetrics '97*, Seattle, WA, pp. 181–191.

Knightly, E. W. and N. B. Shroff (1999, March/April). Admission control for statistical qos: Theory and pratice. *IEEE Network 13*(2), 20–29.

Konstantopoulos, T. and S.-J. Lin (1996, Dec.). High variability vs long-range dependence for network performance. In *Proc. 35th. IEEE CDC*, Kobe, pp. 1354–1359.

Kontovasilis, K., S. Wittevrongel, H. Bruneel, B. V. Houdt, and C. Blondia (2002). Performance of telecommunication systems: Selected topics. In *Communication Systems: the State of the Art*. Kluwer Academic Publishers.

Leland, W., M. Taqqu, W. Willinger, and D. Wilson (1994). On the Self-Similar Nature of Ethernet Traffic. *IEEE/ACM Transactions on Networking 2*, 1–15.

li Zhang, Z., D. Towsley, and J. Kurose (1994). Statistical analysis of generalized processor sharing scheduling displine. In *SigComm'94*, London, England, UK, pp. 68–77.

Likhanov, N. and R. Mazumdar (1998). Cell loss asymptotics in buffers fed with a large number of independent stationary sources. In *IEEE Infocom*, pp. 339–346.

Mandjes, M. and N. Boots (2002). The shape of the loss curve, and the impact of long-range dependence on network performance. Technical report.

Mandjes, M. and N. Boots (2004). The shape of the loss curve, and the impact of long-range dependence on network performance. AEU International Journal on Electronics and Communications. to appear.

Mandjes, M. and J. H. Kim (2001). Large deviations for small buffers: an insensitivity result. *Queuing Systems 37*, 349–362.

Mosteller, F. and J. W. Tukey (1977). *Data Analysis and Regression: A Second Course in Statistics*. Reading, Massachusetts, U.S.A.: Addison-Wesley.

Norros, I. (1994). A Storage Model with Self-similar Input. *Queueing Systems 16*, 387–396.

Paxson, V. (1997). Automated Packet Trace Analysis of TCP Implementations. In *ACM SIGCOMM*, pp. 167–179.

Paxson, V. and S. Floyd (1995). Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking 3*, 226–244.

Presti, F. L., Z.-L. Zhang, J. Kurose, and D. Towsley (1999, August). Source time scale and optimal buffer/bandwidth tradeoff for heterogeneous regulated traffic in a network node. *IEEE/ACM Transactions on Networking 7*(4), 490–501.

Ribeiro, V. J., R. H. Riedi, M. S. Crouse, and R. G. Baraniuk (1999). Simulation of NonGaussian Long-Range-Dependent Traffic Using Wavelets. *ACM SIGMETRICS 27*, 1–12.

Riedi, R. H., M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk (1999). A Multifractal Wavelet Model with Application to Network Traffic. *IEEE Transactions on Information Theory 45*(3), 992–1019.

Shroff, N. B. and M. Schwartz (1998, August). Improved loss rate calculations at an atm multiplexer. *IEEE/ACM Trans. on Networking 6*(4), 411–422.

Simonian, A. and J. Guibert (1995). Large deviations approximation for fluid queues fed by a large number of on/off sources. *IEEE Journal of Selected Area in Communications 13*, 1017–1027.

Stevens, W. R. (1994). *TCP/IP Illustrated, Volume 1: The Protocols*. Addison-Wesley.